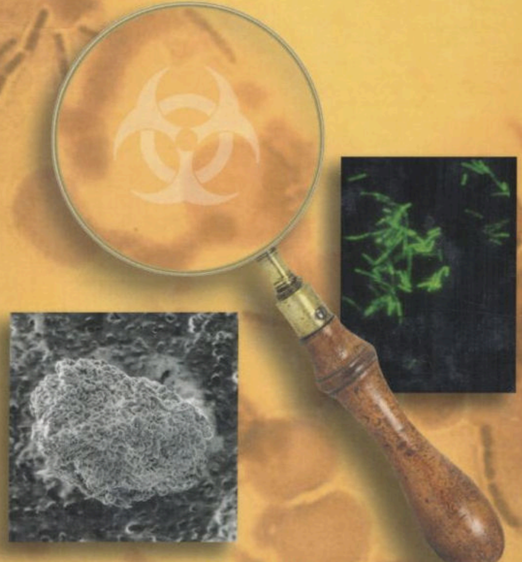
Copyrighted Material

ELSEVIER
ACADEMIC
PRESS

MICROBIAL FORENSICS

Roger G. Breeze
Bruce Budowle
Steven E. Schutzer



Copyrighted Material

VISIT…

LANZAROTE
Caliente.COM

Elsevier Academic Press
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
84 Theobald's Road, London WC1X 8RR, UK

This book is printed on acid-free paper. ♾

Copyright © 2005, Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone: (+44) 1865 843830, fax: (+44) 1865 853333, e-mail: permissions@elsevier.co.uk. You may also complete your request on-line via the Elsevier homepage (http://elsevier.com), by selecting "Customer Support" and then "Obtaining Permissions."

Library of Congress Cataloging-in-Publication Data
Application submitted.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 0-12-088483-6

For all information on all Elsevier Academic Press publications
visit our Web site at www.books.elsevier.com

Printed in the United States of America
05 06 07 08 09 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

Contributors xiii
Foreword xvii
Preface xix
Acknowledgments xxi

1 Microbial Forensics

Bruce Budowle, James P. Burans, Roger G. Breeze, Mark R. Wilson, and Ranajit Chakraborty

Introduction 1
The Threat 2
History 3
Forensic Science 6
Defining the Microbial Forensics Program 8
Carrying Out a Successful Microbial Forensics Program 11
SWGMGF 12
Epidemiology (A First Step in a Microbial Forensic Analysis) 16
Microbial Forensic Tools 17
Bioinformatics and Genetics 19
Forensic Case Examples 20
Education 21
Conclusion 23
Acknowledgment 23
References 23

2 Infectious Diseases: Not Just a Health Matter Anymore

Roger G. Breeze

Introduction 27
Infectious Disease: The Classic Health Viewpoint 29
Infectious Disease: The Offensive Biological Weapons Viewpoint 31
Infectious Disease: New Players and New Threats 32
Infectious Disease: What Are the Threats? 36
Infectious Disease: Biosafety and Biosecurity 37
References 39

3 The Fundamentals of Human Virology

Jeffrey Wilusz

Introduction 41
An Overview of the DNA Viruses 45
An Overview of the RNA Viruses 47
Positive-sense RNA Viruses 47
Negative-sense Single-Stranded RNA Viruses 49
Segmented RNA Viruses and the Retroviridae 50
The Application of Molecular Forensics and Epidemiology to Viral Infections 51
References 53

4 Keeping Track of Viruses

Jack Hietpas, Laura K. McMullan, David P. Mindell, Holly L. Hanson, and Charles M. Rice

Introduction 55
What Is a Virus? Basic Virology 56
Virus Lifecycle 57
Virus Classification 60
Generation of Viral Diversity 62
Introduction to Viral Kinetics and Outcome 63

5 Bacterial Pathogens

Paul Keim

DNA Regions of Value for Forensic Identification of Bacterial Pathogens 99
Bacterial Pathogens of Concern 102
References 106

6 Biology and Detection of Fungal Pathogens of Humans and Plants

Gary A. Payne and Marc A. Cubeta

Introduction 110
Classification, Nomenclature, and Taxonomy of Fungi 111
Growth Habit and Reproduction 115
Pathogenicity 118
Epidemiology 123
Detection and Identification 125
References 128

7 Forensic Aspects of Biologic Toxins

James D. Marks

Introduction 131
Botulinum Neurotoxin and Botulism 132
Ricin 146
Staphylococcal Enterotoxin B 147
Clostridia perfringens Epsilon Toxin 148
Summary 149
References 149

8 Epidemiologic Investigation for Public Health, Biodefense, and Forensic Microbiology

Stephen A. Morse and Ali S. Khan

9 Molecular Epidemiology and Forensics of RNA Viruses

Consuelo Carrillo and Daniel L. Rock

Challenges Posed by RNA Viruses 174
A Case Study: Foot-and-Mouth Disease in the United Kingdom, 2001 175
What Needs To Be Done? 177
References 180

10 Investigation of Suspicious Disease Outbreaks

Lynda Collins Kelley and Roger G. Breeze

Natural and Deliberate Disease 187
Importance of Disease Surveillance 190
Global Disease Reporting Systems 191
U.S. Disease Surveillance and Reporting Systems 192
Epidemiologic Investigation 194
Composition of an Outbreak Investigation Team 194
Epidemiologic Analysis 195
Evaluation of Clinical Findings 200
Investigation of Animal Disease Outbreaks 201
Investigation of Crop Disease Outbreaks 202
Specimen Collection 203
Environmental Sampling 203
Sample Handling 204
Specimen Storage 204
Laboratory Analysis 204
Serology 205
Evaluation of the Etiologic Agent 205
Analysis of Meteorological and Climatic Conditions 206
Economic Analysis 206
Media Reports 207
Available Documents 207
Evaluation of Scientific Literature 207
Training or Work Experience 208
Equipment or Potential Means of Delivery 208

Biological and Physical Integrity of a Sample 218
Legal Concerns for Sample Handling and Data Records 219
Safety Issues 223
Sample Shipping Regulations 225
CDC/NIH Regulations 226
References 230

12 Forensic Genetic Analysis of Microorganisms: Overview of Some Important Technical Concepts and Selected Genetic Typing Methods

Charles L. Cooke Jr.

Introduction 233
Useful Definitions and Concepts 235
Taxonomy and Epidemiology 235
Genetic Considerations 236
Restriction Endonucleases and Polymerase Chain Reaction 237
Of Dendrograms and Phylogenetic Trees 240
Molecular Genetic Techniques for Strain Typing 241
Multilocus Sequence Typing 241
Restriction Fragment Length Polymorphism Typing 242
PCR-Based Genetic Typing 244
Viruses and Fungi 245
Conclusions 246
References 247

13 Non-DNA Methods for Biological Signatures

Charlene M. Schaldach, James J. DeYoreo, Tony Esposito, David P. Fergenson, Eric Gard, Christopher Hollars, Thomas Huser, Stephen M. Lane, Alexander J. Malkin, Maurice Pitesky, Chad Talley, Herb J. Tobias, Bruce Woods, Graham Bench, Patrick Grant, Michaele Kashgarian, John Knezovich, James Ferreira, Kuang-Jen Wu, Joanne Horn, and Stephan P. Velsko

14 Microbial Forensics Host Factors

Steven E. Schutzer

General Concepts 296
Utility of Serologic Analysis of People Exposed to Anthrax: Strengths and Limitations 302
Considerations and Concerns Raised by Analysis of Other Infections 306
Possible Scenarios of Bioterrorism Attacks: Distinguishing Victims from Perpetrators 308
References 311

15 Bioinformatics Methods for Microbial Detection and Forensic Diagnostic Design

Tom R. Slezak and Steven L. Salzberg

A Working Definition of Bioinformatics 313
An Overview of Microbial Diagnostics 314
Detection Diagnostics 315
Nucleic Acid Detection Diagnostics 316
Chemistries for Nucleic Acid Detection 317
Protein Detection Diagnostics 319
Monoclonal Antibody 319
High-Affinity Ligands 320
Forensic Diagnostics 321
Large-Scale, Chip-Based Techniques 323
Protein Forensic Diagnostics 324
Genome Sequencing and Analysis Techniques 324
Basecalling and Accuracy 325
Draft versus Finished 327
Strategies for Bacterial and Viral Sequencing 328
Assembly 328
Annotation 330
Tools for Annotation 330
Annotation for Diagnostics versus Traditional

Tools for Protein Structure Analysis 341
Visualization 341
Other Forensic Techniques 342
Protein Mass Spectrometer Analysis 343
Image Analysis 343
Examples 343
Anthracis Analysis 344
Nucleic Acid Signature Pipeline 345
Determination of Target Pathogen Consensus Sequence 346
Fast, Scalable Sequence Comparison Programs to Locate Unique Sequence 347
Primer and Probe Selection and Acceptance 349
Protein Signature Pipeline 350
Acknowledgments 350
References 350

16 Population Genetics of Bacteria in a Forensic Context

Richard E. Lenski and Paul Keim

Introduction 355
DNA Forensics of Humans and Bacteria 356
Case Study of *Bacillus anthracis* 357
Conclusions 366
Recommendations 367
Acknowledgments 367
References 368

17 Quality Management in Forensics Laboratories

Joseph M. Campos

Introduction 371
Laboratory Quality Management 373
A Few Definitions Are in Order at This Point 373
Laboratory Accreditation 374
Validation of Laboratory Tests 375

Laboratory Security 378
Conclusions 378
Bibliography 379

18 Admissibility Standards for Scientific Evidence

Rockne Harmon

Legal Admissibility 382
Background to the *Daubert* Decision 383
Analysis of the *Daubert* Decision 383
The DNA Admissibility Litigation 386
Fingerprint Admissibility Litigation Under *Daubert* 389
Observations 390
References 391

Appendix

Quality Assurance Guidelines for Laboratories Performing Microbial Forensic Work 393

Produced by the Federal Bureau of Investigation-Hosted Members of the Scientific Working Group on Microbial Genetics and Forensics (SWGMGF)

Subject Index 411

CONTRIBUTORS

Numbers in parentheses indicate the pages on which the author's contribution begins.

GRAHAM BENCH (251), Center for Accelerator Mass Spectrometry, Energy and Environment Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

ROGER G. BREEZE (1, 27), Centaur Science Group, Washington, DC 20007

BRUCE BUDOWLE (1), Federal Bureau of Investigation, Laboratory Division, Quantico, Virginia 22135

JAMES P. BURANS (1), National Bioforensic Analysis Center, Department of Homeland Security, Frederick, Maryland

JOSEPH M. CAMPOS (371), Department of Laboratory Medicine, Children's National Medical Center; Professor of Pediatrics, Pathology, and Microbiology/Tropical Medicine, George Washington University Medical, Washington, DC 20037

CONSUELO CARRILLO (173), Plum Island Animal Disease Center, Agricultural Research Service, U.S. Department of Agriculture, Greenport, New York 11944

RANAJIT CHAKRABORTY (1), Department of Environmental Health, Centre for Genome Information, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267

CHARLES L. COOKE JR. (233), Bristow, VA 20136

MARC A. CUBETA (109), Department of Plant Pathology, North Carolina State University, Raleigh, North Carolina 27695-7567

JAMES J. DEYOREO (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

TONY ESPOSITO (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

JOHN W. EZZELL (213), U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Maryland 21702-5001

DAVID P. FERGENSON (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

JAMES FERREIRA (251), Materials Science and Technology Division, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

ERIC GARD (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

PATRICK GRANT (251), Center for Accelerator Mass Spectrometry, Energy and Environment Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

HOLLY L. HANSON (55), Laboratory of Virology and Infectious Disease, Center for the Study of Hepatitis C, The Rockefeller University, New York, New York 10021

ROCKNE HARMON (381), Alameda County District Attorney's Office, Oakland, California 94612

JACK HIETPAS (55), Laboratory of Virology and Infectious Disease, Center for the Study of Hepatitis C, The Rockefeller University, New York, New York 10021

CHRISTOPHER HOLLARS (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

JOANNE HORN (251), Energy and Environment Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

THOMAS HUSER (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

MICHAELE KASHGARIAN (251), Center for Accelerator Mass Spectrometry, Energy and Environment Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

PAUL KEIM (99, 355), Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona 86011-5640

LYNDA COLLINS KELLEY (187), U.S. Department of Agriculture, Food Safety and Inspection Service, Office of Public Health Science, Russell Research Center, Athens, Georgia 30604

ALI S. KHAN (157), Division of Parasitic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333

JOHN KNEZOVICH (251), Center for Accelerator Mass Spectrometry, Energy and Environment Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

STEPHEN M. LANE (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

RICHARD E. LENSKI (355), Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48825-4320

ALEXANDER J. MALKIN (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

JAMES D. MARKS (131), Department of Anesthesia and Pharmaceutical Chemistry, University of California, San Francisco, San Francisco General Hospital, San Francisco, California 94110

LAURA K. MCMULLAN (55), Laboratory of Virology and Infectious Disease, Center for the Study of Hepatitis C, The Rockefeller University, New York, New York 10021

DAVID P. MINDELL (55), Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, Ann Arbor, Michigan 48109-1079

STEPHEN A. MORSE (157), Bioterrorism and Response Program, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333

GARY A. PAYNE (109), Department of Plant Pathology, North Carolina State University, Raleigh, North Carolina 27695-7567

MAURICE PITESKY (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

CHARLES M. RICE (55), Laboratory of Virology and Infectious Disease, Center for the Study of Hepatitis C, The Rockefeller University, New York, New York 10021

DANIEL L. ROCK (173), Plum Island Animal Disease Center, Agricultural Research Service, U.S. Department of Agriculture, Greenport, New York 11944

STEVEN L. SALZBERG (313), Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland; The Institute for Genomic Research, Rockville, Maryland 20878

CHARLENE M. SCHALDACH (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

STEVEN E. SCHUTZER (295), Department of Medicine, University of Medicine and Dentistry of New Jersey—New Jersey Medical School, Newark, New Jersey 07103

TOM R. SLEZAK (313), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

CHAD TALLEY (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

HERB J. TOBIAS (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

STEPHAN P. VELSKO (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, and Nonproliferation, Arms Control, and International Security Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

MARK R. WILSON (1), Federal Bureau of Investigation, Laboratory Division, Quantico, Virginia 22135

JEFFREY WILUSZ (41), Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado 80523-1619

BRUCE WOODS (251), BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

KUANG-JEAN WU (251), Materials Science and Technology Division, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

FOREWORD

Kilmacthomas is a small town of some 7,500 inhabitants, not far from Waterford in southern Ireland. In November 2001, the citizens of Kilmacthomas found themselves subject to the same fears that swept communities from Washington, D.C., to Nairobi, Karachi, Beijing, and thousands of other cities, towns and villages across the world. The *Waterford News & Star* reported, together with the results of the whist drive in Ballylaneen, that the Kilmacthomas post office had been closed after a white powder was found in the mail box. Police, fire department, ambulance and the Army Disposal Unit personnel were called to the scene to make the post office safe. The powder did not contain anthrax but the incident shows how easily the fear of bioterrorism paralyzes communities.

We have always lived in perilous times. Seemingly from nowhere, "new" or new variants of infectious agents have appeared and afflicted human beings. The Spanish flu of the 1918–1919 pandemic killed tens of millions; in 1984, human immunodeficiency virus (HIV) was identified as the causative agent of a new disease, acquired immunodeficiency disease syndrome (AIDS); hantavirus pulmonary syndrome caused by the Hanta virus of rodents was first identified in 1993; variant Creutzfeldt-Jakob disease was first diagnosed in 1996; and severe acute respiratory syndrome (SARS) caused widespread panic in 2003. This list does not include the devastating infections of agricultural plants and animals—the 2001 foot-and-mouth disease outbreak in the United Kingdom led to the slaughter of 11 million animals.

In 1991, Lyme disease was still a "newly emerged" infection; the bacterium responsible had been identified only nine years earlier. In that year, the first in a series of meetings on Lyme disease was held at the Banbury Center of the Cold Spring Harbor Laboratory. It was chaired by Steve Schutzer. It highlighted many diagnostic problems common to other newly emerging infections. How can one diagnose a previously-unknown infection? What laboratory tests can detect an unknown pathogen? This book has its origins in that series of meetings, which by 2000 had metamorphosed to embrace the wider problems of novel infections in plants and animals, as well as human beings. It was during the 2000 conference on Meeting the Challenge of Infectious Diseases in the 21st Century, chaired by Roger Breeze, that two sessions were devoted to "Detection, Identification, Forensics and Diagnosis." Then came the anthrax terrorist attacks of 2001, when genuine threats were followed by many thousands of hoaxes throughout the world. And so it was that the editors of this book—Steve Schutzer, Roger Breeze, and Bruce Budowle, along with other colleagues—organized Banbury Center meetings in 2002 and 2004 to examine forensic approaches to microbial bioterrorism. This was not the first occasion on which forensics had been discussed at Banbury. In 1988, there was what turned out to be an historic meeting examining the early implementation of human DNA fingerprinting. Bruce Budowle was one of the participants in that meeting.

This book tackles many of the issues facing investigators of real or potential attacks employing microorganisms and toxins. How to identify rapidly unknown substances? If there is an outbreak or even just one example of a seemingly new disease, is it naturally occurring? If not, who is responsible? What pathogen is involved? Has it been manipulated in any way? What advice should be given to the officials and to the public? *Microbial Forensics* takes us from the fundamental biology of pathogenic organisms, through the investigation of suspicious events, to the legal requirements for prosecution should a felony have been committed. Such investigations require expertise in many different fields and the authors of the chapters in the book are acknowledged experts, bringing their knowledge and experience to bear on what continues to be one of the most difficult challenges of the new, post September 11th world. *Microbial Forensics* provides a much needed resource for all those faced with investigating mysterious white powders and unknown infections.

Jan Witkowski
Director, Banbury Center
Cold Spring Harbor Laboratory
New York

PREFACE

Pathogenic microbes and their toxins have always posed a significant threat to the health of humans, animals and plants that become exposed and infected in the great scheme of Life. But now there is a new threat—the deliberate use of pathogens and toxins as weapons in acts of bioterrorism or the commission of biocrimes. Countering these weapons demands new tools and a new scientific discipline, microbial forensics. We define microbial forensics as a scientific discipline dedicated to analyzing evidence from a bioterrorism act, biocrime, or inadvertent microorganism/toxin release for attribution purposes.

Epidemiologists, particularly those in public health fields, have used forensic practices for decades to identify causative agents and the etiology of disease. Some essential elements of this new discipline aimed at attribution have long been growing in several government agencies. But microbial forensics could not evolve into its own discipline because it was not considered a potential deterrent and it seemed limited in its ability to attribute the source or to identify the perpetrator. However, over the last decade, and particularly after the 2001 anthrax letter attacks, microbial forensics has crystallized and become more formalized. This has brought renewed interest, an influx of resources and people, and new rules of engagement.

Scientists entering the field of microbial forensics and people who may make political, administrative, or legal decisions or just would like to be better informed have no single source to consult. This book was written to provide such a resource, to introduce the discipline, and to describe some of the challenges and opportunities ahead. Not all possible topics are described. Instead, representative examples are provided to initiate and educate the reader. We hope that the book will serve as a foundation text to stimulate new generations of scientists and legal experts who will bring additional expertise to the field. The book has this Preface and 18 other chapters that address: 1) definition of the discipline, select history of past bioterrorism and biocrime events, challenges and direction; 2) synopses of the basic biology of viruses, bacteria, fungi, and toxins; 3) epidemiology; 4) laboratory and technology considerations; 5) host-pathogen interactions; 6) bioinformatics and population genetics; 7) quality assurance; and 8) legal aspects. An appendix lists quality assurance guidelines for microbial forensic laboratories that carry out analytical assays.

Over the past century, science has played an increasingly greater role in criminal investigation. Microbial forensic science will continue this tradition and assist in solving crimes by characterizing physical evidence found at a crime scene for attribution purposes. A combination of diverse disciplines is exploited to analyze evidence, including biology, microbiology, medicine, chemistry, physics, statistics, population genetics, and computer science. New techniques must be employed to get the most evidence from biological materials used in terrorist and criminal events, especially when more traditional forms of evidence are either not available or very limited in content. These concepts are explained in this book. Yet scientific analysis alone is not enough. Forensic science should be integrated with other information. In practice, rigorous attention must be paid to steps that will assure admissibility of results in a court of law. Admissibility rules, and legal expectations and consequences, are described by citing past cases. We also stress quality control and quality assurance as the means to ensure reliability. Practices such as adherence to chain of custody procedures, documentation of activities, using tested reagents, calibrated equipment, negative and known positive control samples, validated procedures, standard operating procedures, and so on, are the essence of reliability and confidence. These in turn ensure admissibility.

The foundations described here should be strengthened, built upon, and when appropriate remodeled. The driving force for the maturation of microbial forensics will be our present and future colleagues. We look forward to their input, interaction, and insight.

Bruce Budowle
Steven E. Schutzer
Roger G. Breeze

ACKNOWLEDGMENTS

This book grew out of a meeting at Cold Spring Harbor Laboratory attended by some 40 scientists who were enthusiastic about sharing their experiences and ideas and anxious to see the new discipline of microbial forensics described in a foundational reference text. But any editor who has ever awoken the morning after such group excitement knows that euphoria soon fades, to be replaced by hard realities and the relentless demands of other commitments. In this case, the work could be shared equally among three editors who brought different talents to the whole enterprise. However, two of us, Roger Breeze and Bruce Budowle, would like to take this opportunity to recognize the insight, tireless enthusiasm, optimism, discipline and leadership of Steve Schutzer, without whom this volume would never have seen the light of day. Steve was instrumental in putting the Cold Spring Harbor Laboratory-Banbury Center meetings together, stimulating discussions during the formal meetings and evening sessions, keeping manuscripts on track, and ensuring the other editors met their obligations.

Two people played vital roles in establishing microbial forensics as a discipline through vision, leadership, advocacy and pioneering accomplishments from the early years: Janet Dorigan of the Office of Research and Development, Central Intelligence Agency, and Randall Murch of the Laboratory Division of the Federal Bureau of Investigation. On behalf of all the authors in this volume, we would like to recognize their contributions and express our appreciation. We extend the same sentiments and recognition to others at the same agencies who have contributed substantially to the foundations of microbial forensics but remain anonymous.

We also thank the contributors to this book, those who provided intellectual input at Cold Spring Harbor Laboratory and Banbury Conference meetings over the years, and the many others whose efforts have contributed to the foundation of microbial forensics. Government agencies whose support has been invaluable include: the Department of Defense, the Intelligence community, the Department of Justice and National Institute of Justice, the Department of Homeland Security, the Department of Agriculture, the Department of Energy, the Centers for Disease Control and Prevention, the National Institutes of Health, and the Food and Drug Administration. The National Academy of Sciences has contributed to this field and should be commended.

Bruce Budowle, on behalf of the Federal Bureau of Investigation, would like to thank the members of the Scientific Working Group on Microbial Genetics and Forensics for their contributions and continued efforts. He also would like to acknowledge the efforts of his colleagues at the Chem-Bio Sciences Unit at the Federal Bureau of Investigation, and the encouragement of his close friends Arthur J. Eisenberg, Leonard Klevan, Jose A. Lorente, Randy Murch, Antti Sajantila, and Rock Harmon.

Roger Breeze would like to acknowledge Hugh Pirie, Norman Whittle, David C. Taylor, Mick Taylor, Leo Bustad, and Robert Wilson without whose support and encouragement he would not have come to this day.

Steven Schutzer would like to thank friends, family, and colleagues. He particularly would like to thank Bruce Budowle and Roger Breeze for their vision and the privilege of working with them. Special thanks to Jan Witkowski and Jim Watson for providing the unique environment that helps launch new endeavors.

Steven E. Schutzer
Roger G. Breeze
Bruce Budowle

CHAPTER 1

Microbial Forensics

BRUCE BUDOWLE AND MARK R. WILSON
Federal Bureau of Investigation, Laboratory Division, Quantico, Virginia

JAMES P. BURANS
National Bioforensic Analysis Center, Department of Homeland Security, Frederick, Maryland

ROGER G. BREEZE
Centaur Science Group, Washington, DC

RANAJIT CHAKRABORTY
Center for Genome Information, University of Cincinnati College of Medicine, Cincinnati, Ohio

INTRODUCTION

Officials, victims, and family and friends of victims of crime often want to identify the perpetrator of the crime and to exclude those who could not have committed the crime. Science can assist in that endeavor by characterizing physical evidence found at the crime scene. The fields of human DNA analysis, hair morphology comparisons, handwriting, fingerprint identification, drug analysis, ballistics, tool marks, and others are well established in forensic science.

In the 21st century, the ability to manipulate and disseminate pathogens has increased to a point where there are grave concerns about potential use of microbiological organisms and their toxins to attack and cause serious harm to humans, animals, and plants. These new scenarios of threat and crime can also be addressed using forensic science practices for identifying perpetrators of crimes and to protect the innocent. The field of microbial forensics, with the goal of attribution, has begun to crystallize, and much effort nationally and internationally is being dedicated to develop and form this field into a mature science. This introductory chapter describes the efforts and needs of the microbial forensics field. First, the threat is defined. Then, historical examples of bioweapons use are provided not only to familiarize readers with the potential use, but to show that such weapons do exist and have been used. As

Microbial Forensics
Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

some readers may not be familiar with forensic science, a general description of the basics of forensic science practices are provided in a following section. Subsequently, the microbial forensic field is defined, and the efforts and needs of the field are discussed. Lastly, education efforts are stressed as a key to disease surveillance and appropriate public response to a disease outbreak, whether it is intentional or natural.

THE THREAT

Terrorism can be defined as an attack or threat of an attack on the innocent to create fear, intimidate, inflict harm, and/or affect economic well-being. These acts have often been politically motivated, but may not always be so. To create terror, perpetrators of bioterrorism use pathogens or toxins as weapons. A number of bacteria, viruses, and fungi pose serious health risks to humans, animals, and plants, and the use of them as bioweapons can have serious consequences on human health, economic development, social stability, and political activities of nations worldwide. Moreover, technology has developed to such a point that there is a greater potential and increased likelihood that biological weapons will be accessible to individuals or small groups of individuals instead of just state organized institutions. Sophisticated high technology, such as that used for nuclear weapon development, is not needed to produce large quantities of weaponizable pathogens, and large quantities of bioweapons can be easily produced. Dissemination of pathogenic agents has been considered difficult in the past and put a limitation on their use, but some approaches can be relatively simple, such as may be needed to infect large herds of livestock. To add to the motivation of use is the low cost of producing and using pathogenic agents in criminal acts. It has been estimated that the cost to inflict civilian casualties is about \$2000 per km^2 for conventional weapons and about \$800 per km^2 for nuclear weapons; for biological weapons the cost is only about \$1 per km^2.[1,2] Thus, bioterrorism will be one of the major threat challenges of the 21st century. Indeed, the anthrax letter bioterrorism attack of 2001 brought to the forefront the need to consider strengthening homeland security and enhance our forensic capabilities for attribution and deterrence.

In addition to bioterrorism, pathogens and/or toxins may be used in biocrimes. These acts may be considered the same as traditional crimes which usually are directed towards harming individuals except that the weapon is biological in nature, instead of guns, knives, and traditional chemical poisons such as cyanide. However, the use of a bioweapon in a criminal case should be taken as seriously as that of a bioterrorist attack.

HISTORY

The use of pathogenic agents as weapons has been documented for more than two millennia. The ancient Romans carried out biological warfare by putting carrion into the wells of their enemies to poison the water supply. Such crude biowarfare tactics continued well into the 20th century.[3,4] During the 14th-century siege of Kaffa (now Feodossia, Ukraine), the attacking Tartar soldiers experienced an epidemic of bubonic plague. The Tartars capitalized on the devastating disease and used the bodies of plague victims as weapons. They catapulted their own soldiers' diseased bodies into the walled city of Kaffa in an attempt to inflict harm on their enemy and likely contributed to the European plague epidemic during the Middle Ages.[5–7] Smallpox and measles contained within blankets and clothing were used by the Conquistadors and British as biological weapons against Native Americans. It is believed that such tactics influenced the outcome of the French and Indian Wars.[5,7] Plague was used as a weapon by the Japanese during World War II. Laboratory-raised fleas were allowed to feed on plague-infected rats. The fleas were then disseminated from airplanes over China.[7,8] In 1993, the Aum Shinrikyo Cult (known for its sarin nerve gas attack in a Tokyo subway) attempted to spread anthrax in Tokyo from the rooftop of an eight-story building owned by the Cult. The purported attack did not cause any illness from the anthrax agent, because the strain Sterne (a vaccine strain) was used. It is believed that the Aum Shinrikyo Cult also dispersed botulinum toxin, Q fever, and spores from a poisonous mushroom, but all attacks failed.[9–11] In 1984, cult followers of Baghwan Sri Rajneesh attempted to affect the outcome of a local election in Dalles, Oregon. They attempted to incapacitate the population and prevent them from voting by successfully contaminating salad bars in 10 restaurants with *Salmonella typhimurium*; 751 people developed food poisoning. A *Salmonella typhimurium* strain found in a laboratory at the commune of the religious cult was indistinguishable from the strain from the outbreak.[12] In 1996 in Dallas, Texas, there was a documented case of intentional Shigellosis contamination. Twelve people of the laboratory staff of St. Paul Medical Center hospital developed severe, acute diarrheal illness. *Shigella dysenteriae* type 2 was isolated from stool samples of some of the sick workers. All outbreak patients reported eating muffins or doughnuts in a break room. *Shigella dysenteriae* type 2, recovered from the patients, from an uneaten muffin, and from the medical laboratory's stock strain were indistinguishable by pulsed-field gel electrophoresis. A criminal investigation ensued focusing on a laboratory technician who was subsequently convicted on five felony assaults and falsifying laboratory documents.[10,13]

The few examples described above focus mostly on those agents that could cause harm to human targets (see CDC list).[5] However, disruption of food sources and great economic upheaval could arise by targeting agriculture. U.S. commercial animal agriculture revolves around three commodities: cattle, swine, and poultry (i.e., chickens and turkeys). Sheep, goats, ducks, and other species may be considered commercially insignificant, but an attack on these could hurt smaller economic sectors and affect public confidence. While horses are not considered to be a domestic livestock industry by the United States Department of Agriculture (USDA), they do fulfill personal and recreational needs. Deer, elk, bison, and other wildlife species are also not considered domestic livestock but could become reservoirs for pathogens, and attacks on wildlife could affect the ecosystem. Fish and shellfish, which form a notable industry, are often not thought of as targets, but could be targets for an attack.

International lists of livestock diseases that pose the greatest threats are maintained by the International Organization of Epizootics (OIE).[14] There are about 15 diseases on the OIE List A and a much larger number in List B. Bluetongue, Newcastle disease, and vesicular stomatitis, which are on List A, occur naturally in the U.S. Terrorists and criminals have easier access to these and other animal pathogens, because they occur naturally. Some animal diseases are caused by zoonotic pathogens (i.e., causing disease in animals and humans). These include: Rift Valley Fever (*Bunyaviridae*), Glanders (*Burkholderia mallei*), and Q fever (*Coxiella burnetii*). Pathogens that reside in a broad range of host species can be more difficult to eradicate and enable access from a variety of reservoirs.

Livestock are highly vulnerable. One need only consider the most recent outbreak of Foot and Mouth disease in England to appreciate the impact of just a natural outbreak. The same or greater effect can be imagined with the use of a livestock-directed bioweapon. Animal pathogens are often highly contagious and can be dispersed without advanced technology; in addition, disease can spread rapidly, animal to animal. In the U.S., disease spread can be exacerbated because livestock are often transported in large numbers throughout the U.S.[15] Historically, attacks have occurred against livestock. For example, during World War I, Glanders (*Burkholderia mallei*) was used by the Germans to infect Allied horses and mules. In fact, a laboratory was established in Chevy Chase, Maryland in 1916 to culture Glanders (and anthrax) and the live organisms were then provided to German agents to infect U.S. draft animals. They successfully infected a number of animals in the U.S., as well as in Europe.[7,10,16] The former Soviet Union also used Glanders in the early 1980s during its war in Afghanistan.[17,18] In 1952, the Mau Mau poisoned and killed some cattle at a British mission station in Kenya using a local toxic plant, the African milk bush (*Synadenium compactum* "Ruby"). They purportedly placed plant latex material into incisions made in the animals' skin.[10,17,18]

There are no major plant pathogens that also pose a significant threat to human public health. However, the plant agriculture industry is a large part of the U.S. economy and a major source of foodstuffs. A significant proportion of the global crops of soybeans, wheat, and corn are grown in the U.S. A successful attack on any of these commodities would have obvious economic impact locally, if not globally, and could affect food supplies. In addition, many other crops such as sorghum, rice, barley, cotton, and tree crops such as citrus, apples, and stone fruits are economically important in regions of the U.S. An attack on a regional crop, such as rice in the U.S., may not be catastrophic economically or significantly affect food supplies, but public concern would intensify. However, an attack on the rice crop in some Asian countries, such as India or China, could be economically devastating and/or create a famine. Also, one needs to consider forest plant life as a significant target that can impact on the ecosystem, economy (e.g., lumber), recreation, and security.

The most significant plant pathogen threats for U.S. agriculture are fungi[19] such as soybean rust (*Phakospora pachyrhizi*)[20,21] and wheat stem rust (*Puccinia graminis*).[22] Wheat stem rust can be an especially effective bioweapon because it can remain viable for more than two years with proper storage conditions and spreads rapidly.[23] Wheat smuts are endemic around the world, and large infestations can greatly reduce crop yield. In addition, wheat smuts may have another weaponizable feature not typically considered; they produce trimethylamine gas that is highly flammable. Harvesters containing infected wheat could perhaps become bombs.[23] Although chemicals such as Agent Orange for defoliation or chlordane to poison food[17] have been used, there are no documented cases of intentional use of pathogens to attack crops or other plants. Natural outbreaks, however, can provide insight into the possible effect of a successful attack. The potato blight in Ireland during 1845–1846 resulted in a famine that caused one million deaths and the emigration of another million Irish.[23,24] Also in the 19th century, a coffee leaf rust destroyed plantations throughout southeast Asia.[23,24] Brown spot rice contributed to the Bengal famine of 1943.[24] In 1970 in the United States, a leaf blight destroyed one billion dollars worth of corn.[23] Some of the most devastating crop pathogens pass from plant to plant in an airborne manner or by splashing rain; thus, once an infestation has begun, dispersal can occur easily. There are more than 50,000 known diseases that affect crops. Many crops have low genetic diversity, which render some crop varieties widely susceptible to the terrorist or criminal use of plant pathogens. There are few restrictions on trade and transportation of plant products, especially seeds; thus, an attack on agriculture could be easy to carry out and difficult to prevent. Because fungi are far more complex than viruses, bacteria, or toxins (the major human and animal threats), a greater resource investment will be required to develop the investigative tools and technologies required for forensic detection and analysis.

Toxins, produced by some plants, fungi, and bacteria, also may be used as weapons. Botulinum toxin is one of the most lethal toxins known.[5] A single gram of the toxin is sufficient to kill more than 1 million people.[25] Moreover, as a potential weapon, botulinum toxin is easy to produce and transport. At least three times in the early to mid-1990s, the Aum Shinrikyo attempted unsuccessfully to disperse botulinum toxin in Tokyo and at U.S. military installations in Japan.[25] The castor bean (*Ricinus communis*) contains the toxin ricin, which is highly lethal can be made easily (recipes can be readily found on the internet) has been involved in several U.S. criminal cases, and has been used as a tool for assassination.[7] In London in 1978, Georgi Markov, a Bulgarian dissident, was assassinated and ricin was the weapon. He was stabbed in the back of his right leg with the tip of an umbrella, resulting in the deposition of a ricin-laden platinum/iridium spherical pellet only 1.52 mm in diameter. Interestingly, a similar attempt at assassination was made in Paris in 1978 on Vladimir Kostov, a Bulgarian defector. The pellet was lodged in his back, but the delivery mode was unknown. Because Kostov was aware of the death of Markov, the pellet was identified and surgically removed before any lasting effects developed. Ricin was confirmed as the toxin by the presence of ricin antibodies in Kostov.[10] Mycotoxins (toxins from fungi) can cause severe illness and death.[19,26] Even as far back as the 6th century B.C., the Assyrians recognized the potential of mycotoxins as a weapon; they poisoned their enemies' water supply with rye ergot.[26] Shortly after World War II the potential of mycotoxins as a bioweapon became apparent via the Russian military, when *Fusarium*-contaminated flour was used to make bread that was then fed to unaware civilians, who suffered severe illness.[26] Mycotoxins, in the form of yellow rain, allegedly have been used in Laos (1975–1981), Kampuchea (1979–1981), and Afghanistan (1979–1981). As many as 10,000 people, civilians and guerilla fighters, may have died from the mycotoxin poisoning.[26] The Iraqis had an active program developing aflatoxin (although it is difficult to see how this mycotoxin would be used as a significant weapon).[27]

One can appreciate that there are many potential targets, and the number and variety of potential pathogenic weapons are large. Some pathogens are readily available and may be easy to weaponize and disperse. It is now time to build an effective forensic science program to analyze evidence found at crime scenes where bioweapons have been used.

FORENSIC SCIENCE

Forensic science generally is the application of science in the investigation of legal matters.[28] In other words, scientific knowledge and technology are used to serve as witnesses in both civil and criminal matters. While science may not offer definitive solutions to the problems of society, it does provide a special investigative role, particularly in the criminal justice system. The areas of science that are exploited are diverse, but typically include the major disciplines of biology, chemistry, physics, and geology. Within each discipline are many subcategories of science that may be used in a forensic science investigation. For example, within biology are the disciplines of medicine, pathology, molecular biology, immunology, odontology, serology, psychology, and entomology. The specific discipline(s) applied depends on the circumstances of the crime. Mathematics, especially statistics, are used to place weight or significance on observations or data retrieved from crime scene evidence.

The ultimate question usually is “who committed the crime?” (i.e., attribution), and crime scene evidence can play a role in answering that question. Evidence can be any material, physical, or electronic finding that can associate or exclude individuals, victim and/or suspect, with a crime. It typically comprises materials specific to the crime and to control samples for background information. Types of evidence may be fingerprints, blood, semen, saliva, hair, fibers, documents, photos, computer files, videos, firearms, glass, metals, plastics, paint, powders, explosives, tool marks, and soil.

Once a crime scene is recognized and living victims are cared for, the crime scene is secured to preserve the quality of the evidence. Then, forensic science and its practitioners take over by first recognizing and properly collecting and preserving physical evidence. Such processes commence by setting up a documented chain of custody on all evidence collected and by recording the crime scene. These actions are important for scientific and legal integrity. Chain of custody is the practice of maintaining the continuity of possession of evidence and minimizing and/or documenting all those people who have come into contact with the evidence. Importantly, if one does not properly collect and preserve pertinent evidence, crucial and reliable forensic information cannot be obtained. Improper preservation methods may lead to loss of crucial information. Thus, defining proper evidence collection and preservation procedures and the training of those individuals who carry out such activities (including first responders) are fundamental to any forensic analysis process.

Once collected and preserved, the evidence is sent to the laboratory for analysis and all chain of custody practices continue to be maintained. Within the laboratory, a cadre of analytical procedures may be used to identify the material. Often, a comparison is made of the data obtained from the evidence (for example, a DNA profile from a bloodstain) with that obtained from analysis of a reference sample (for example, a DNA profile from a buccal swab from a suspect in a murder case). Ideally, unique identification of the sample, and thus its ultimate source or origin, is desired. In many cases that may not be possible. Thus, one categorizes evidence into two types of characteristics: individual and class. When evidence can be associated with a reference, or source sample with a high degree of probability, it is considered to possess individual characteristics. Typical types of evidence that fall into the individual characteristics category are fingerprints and the current battery of human DNA markers. However, most evidence cannot be characterized to such a level. It cannot be associated with a high degree of confidence with an alleged source sample. Thus, the evidence can only be included within a group and hence is classified as class-characteristic evidence. There is nothing inherently wrong with class-characteristic evidence; it adds direct and/or circumstantial evidence for the fact finder to consider in the totality of the case. At times, it may be difficult to place a probability on class evidence. Yet, a failure to exclude is still valuable information. It is important, though, to recognize and convey the limited value that the analysis may comprise. Examples of evidence that fall into the class-characteristic category are ABO blood groups, mitochondrial DNA, hairs, glass, fibers, and paint. In general, while the class characteristics do not achieve the level of source attribution, they generally offer valuable information with regard to narrowing down the possible source of origin of specimens collected from the crime scene.

Interpretation of data in a comparison of an evidence sample and a reference sample is routinely carried out. Generally, there are three categories of interpretation: inclusion, exclusion, and inconclusive. An inclusion, or match, is stated when the pattern or profile from the two samples is sufficiently similar and potentially could have originated from the same source. An alternate definition of an inclusion is a failure to exclude the two samples as having a common origin or belonging to the same group. An exclusion is stated when the pattern or profile is sufficiently dissimilar such that the two samples could not have originated from the same source. Lastly, an inconclusive interpretation is rendered when there are insufficient data to provide a conclusive interpretation. Obviously, the higher the resolving power that a procedure has, the better is the chance of excluding those wrongly associated with the evidence. Note that we did not say “associated with a crime.” Scientific evidence in itself does not prove guilt or innocence. It is the entirety of all evidence (science and non-science) that is used by the fact finder to determine guilt or innocence.

DEFINING THE MICROBIAL FORENSICS PROGRAM

Most would agree that based on past history and with current technology capabilities, the potential of biological weapons being used is greater than at any other time in history. Only a few individuals are needed with expertise and access to dual-use equipment (e.g., equipment used in the pharmaceutical or food industries) to inexpensively produce a number of bioweapons. As reviewed earlier, microorganisms and their toxins make particularly dangerous weapons because bioweapons can be grown from a single organism or cell; thus creation of large quantities can be accomplished readily and at low cost, starting with minimal amounts of starting material. Thus, even the smallest amounts of pathogens can be used to perpetrate a major terrorist event or biocrime. Furthermore, it can be difficult if not impossible to determine whether very small amounts of such materials have been removed surreptitiously from a facility that is authorized to possess pathogens. Most importantly, a biological weapon can be easy to conceal and, depending on the target, easy to disseminate. We must accept that it is impossible to guarantee that the government can deter or prevent all bioterrorist or criminal acts. Therefore, a proactive approach must be taken. In addition to the physical security measures being enacted, science can be effective as a deterrent and for identifying perpetrators. One major scientific approach to improve investigative capabilities for attribution and deter the use of pathogenic agents in an illicit manner is the development of a strong, scientifically rigorous microbial forensics program.

Microbial forensics can be defined as a scientific discipline dedicated to analyzing evidence from a bioterrorism act, biocrime, or inadvertent microorganism/toxin release for attribution purposes. One may consider attribution solely to be the “fingerprinting” of a pathogenic agent, but the unique identification of a microorganism may never be possible, because of the clonal nature of many microorganisms and, on a case-by-case basis, lack of population and phylogenetic data. More importantly, the ultimate goal of attribution is identification of the persons who committed the bioterrorist act or biocrime, intentionally or inadvertently. In addition to microbiological analytical tools, traditional forensic analyses, such as human DNA analysis, dermatoglyphic patterns, analytical chemistry, tool marks, and other techniques will be used to analyze a bioterrorist event or biocrime evidence.

There is nothing exceptionally unique about forensic microbiological evidence compared with other forensic evidence. Recognizing a crime scene, preserving a crime scene, chain of custody practices, evidence collection and handling, evidence shipping, analysis of evidence, and interpretation of results are carried out in the same general manner as for any forensic evidence, except that the evidence will be handled as a biohazard [even more so than with human immunodeficiency virus (HIV)-infected blood]. It is anticipated that the majority of microbial forensic evidence will fall into the class-characteristic category, with some of the data being very informative and some being rather limited. For example, the bacterium *Bacillus anthracis* demonstrates low diversity among the various strains of the species. If analyzed in the context of a biocrime, one may only be able to identify the strain, while substrain resolution may be difficult. If such data can be derived, for example, by development of a specified biomarker assay, then further resolution may be possible. In contrast, HIV is a rapidly evolving virus. It is likely that two samples with a recent common origin could differ at a few nucleotide bases within the genome. Not having an identical nucleic acid sequence between an isolate and a reference sample in itself is not considered exculpatory evidence. Phylogenetic trees/networks are used to demonstrate relationships of those isolates that have a close or similar ancestry compared with those isolates that have less related histories (see the Forensic Case Examples section below).

Since an attack may be easy to carry out and difficult to prevent, law enforcement needs to enhance its capabilities to confront this new challenge. Law enforcement already has the infrastructure (i.e., its traditional role) for attribution and deterrence. Measures need to be developed and implemented to effectively carry out law enforcement's responsibility in the field of microbial forensics. The major components of a successful microbial forensics program may be listed as:

1. *Detection and Identification* are keys to thwart bioterrorism. To effectively carry out attribution, robust analytical techniques need to be developed and implemented. Assay development to enhance sensitivity and specificity and expand detection capabilities must be promoted. Analytical solutions may have to be developed quickly and effectively. These include DNA-based systems, as well as analytical chemistry and physical analyses (i.e., nonbiological evidence characterization), culture, immunoassays, and the use of bioassays in animals and tissue culture.

2. *Information* and *Databases* will play an important role in the microbial forensics endeavor. The quality and accessibility of rapidly expanding and evolving databases, such as those that contain bioagent genomic sequence data, need to improve. To achieve this, national databases on pathogen genetics and other biological data (and to include nonbiological evidence) need to be created.

There needs to be a relational database on those who have access to these pathogens so that threats can be deterred or traced back effectively to possible sources. Security measures are already being enhanced to restrict and control access to select pathogens and toxins.[29] While such a database may deter some individuals from participating in microbiology research, the infrastructure will more likely protect the legitimate user so that exchange of scientific information can proceed for the betterment of society (such as developing therapeutics and better diagnostic assays). Another database needed is one that is encyclopedic in nature. There are many sources that contain scientific information, including but not limited to publications, presentations, websites, and genomic databases. It is difficult to access all these sites in an effective and rapid manner. Being able to place at one's fingertips all microbiology data and data on associated nonmicrobial forensic materials will greatly enhance the investigative capabilities of the microbial forensic scientist.

3. *A strain repository* to house pathogens and other appropriate near-neighbor microorganisms must be developed. The "near-neighbor" concept is intimately related with the methods used for detection and identification. Some methods of identification that are robust may lead to a broad class of near neighbors, while sophisticated methods may define near neighbors that are narrower. Well-characterized samples must be available to enable good-quality assay development. Assays cannot be adequately validated without proper samples and reference material. Bioinformatic interpretation of analytical results from evidence samples may be more limited without properly defined samples and controls. In addition, better control of access and dissemination of select agents for research and development can be executed.

4. New analytical methods and some existing methods need to be properly *validated*. Such validation tasks are not limited to laboratory procedures; they equally apply to tools to be used for data interpretation and statistical assessment issues. Moreover, some biological crimes may require analysis by methodologies that may not have undergone the rigorous review process of that of standard operating protocols. A preliminary review process for such assays must be implemented (see Quality Assurance discussion below).

5. *Quality assurance* (QA) guidelines for microbial forensic laboratories must be established (some are already enacted due to public health regulations). One must employ high-quality practices to ensure that reliable results are obtained and to maintain public confidence.

CARRYING OUT A SUCCESSFUL MICROBIAL FORENSICS PROGRAM

Building a robust microbial forensics program is a challenging task; there are many aspects to consider, and it is a continuously evolving process. Regardless, dedicating resources and establishing centers of excellence are required. In the U.S., a National Bioforensics Analysis Center (BFAC) is being developed as part of the National Biodefense Analysis and Countermeasures Center (NBACC) and the Fort Detrick, Frederick, Maryland, interagency biodefense campus. The BFAC (and partner laboratory network) will serve as the national reference center to support homeland and national security for the attribution of the use of biological weapons. The laboratory will be part of the NBACC and will be supported primarily by the Department of Homeland Security (DHS) in partnership with the FBI and will execute and coordinate microbial forensic casework. This center will also develop requirements, as well as coordinate and conduct applied development and validation of microbial forensic technologies and methods. In order to carry out its mission, the BFAC will rely on three major components. The first is a Knowledge Center composed of databases on genomics, microbiology, forensics methods, matrix and related evidence assays, bioinformatics, and standardized tools. Readiness of the laboratory can be achieved only by national cooperative efforts. Therefore, the second component is a strong partnership laboratory network with existing government, academic, and private sector assets to include Plum Island, the Department of Defense, the Department of Energy, the Department of Health and Human Services, the National Science Foundation, specialty technology laboratories, and other centers of excellence. The third component is the Scientific Working Group on Microbial Genetics and Forensics (SWGMGF), which will establish and sustain guidelines and/or standards for quality systems, identify processes and procedures, define criteria for knowledge systems, and most importantly serve as an experienced resource on pertinent issues as they arise. The organization of the SWGMGF component is prompted by the success of a similar group (Scientific Working Group of DNA Analysis Methods, SWGDAM) that has been performing similar tasks in the field of human DNA forensics.

SWGMGF

Since the first two components that the national microbial forensic laboratory will rely upon are addressed in other chapters, the focus here will be on the SWGMGF. The newly formed SWGMGF is hosted by the FBI and has a membership drawn from a wide range of federal agencies, as well as academia. Input from the private sector will be sought on a routine basis. As mentioned above, the SWGMGF is modeled after the successful practices in the human DNA forensic arena (i.e., SWGDAM)[30] which utilizes peer consensus to identify and define the criteria on various topics, address technical issues, and set performance guidelines. The Scientific Working Group's success is based on bringing together experienced individuals and organizations (that routinely do not have the opportunity to share and exchange ideas) to foster high-quality, integrated interactions so that challenges related to bioterrorism and biocrimes can be rapidly and effectively addressed. The goals of the SWGMGF are to provide guidance and criteria so that physical evidence can be used to obtain information about the organism or toxin, the persons involved, the places, the processes, the instrumentation, and/or the time of the criminal act.

The missions of SWGMGF are (1) to define criteria for the development and validation of forensic methods for attribution of biological toxins and microbial agents, and (2) to define the needs and criteria for forensic infrastructure and capabilities to support investigation and attribution.

There are many tasks for the SWGMGF to undertake to develop a microbial forensics program and support the mission of the national microbial forensic laboratory. Some may take many years to complete. Because of foundational needs and current funding resource focus, the SWGMGF initially is concentrating on: (1) establishing the criteria for the development and validation of methods to type or individualize various threat agents in ways that can be used forensically to attribute criminal acts, (2) defining quality assurance guidelines, (3) prioritizing research efforts on pathogens and toxins that would most likely be used in biocrimes, (4) understanding and/or enhancing population genetic data so that the significance of a finding is appropriately conveyed, and (5) establishing the design criteria for the various supporting databases. Other topics to be subsequently addressed by SWGMGF include analyses that can further resolve the source of the evidence, such as nonbiological evidence, biological background information (to include pollen, fungi), specific technology needs (to include immunoassays, isotope analysis, mass spectrometry), field testing assays, and host response (such as antibody detection assays).

A primary and immediate focus of the SWGMGF is laboratory quality assurance (QA). Assays carried out in the laboratory need standards of performance for forensic applications. False positive and false negative outcomes and their implications should be understood. Sensitivity and specificity standards should be established. Defined controls and proficiency tests need to be developed. The BFAC will require a valid QA program for operations and to set the standard of performance for the microbial forensic community. Laboratory QA is the documented verification that proper procedures have been carried out by skilled and highly trained personnel in such a way that valid and reliable results can be obtained. When the same methods are used, test reliability implies reproducibility under defined conditions of use and should transcend different laboratories and practitioners. Quality assurance must encompass all significant aspects of the analytical typing process, including organization, management, personnel education and training, facilities, security, documentation of records, data analysis, quality control of reagents and equipment, technical controls, validation, proficiency testing, reporting of results, auditing of the laboratory procedures, and safety. Quality assurance and appropriate guidelines or standards for microbial forensics will be based on human forensic DNA typing standards,[31] clinical laboratories' standards,[32] and standards of the ISO (International Standards Organization).[33] It is important to note that QA guidelines in microbial forensics must retain enough flexibility to accommodate the nature of forensic samples as well as future advancements in technology and molecular biology. The first microbial forensic laboratory QA guidelines document has been published.[34]

One of the important aspects of the QA program is to develop a set of criteria for the development and validation of methods to type various threat agents in ways that can be used to attribute criminal acts. Validation is a process by which a procedure is evaluated to determine its efficacy and reliability and to determine the operational limits of the technique. Many methods are being developed and need to be developed further (many with government support). Some methods, however, may not have been subjected to quality standards, and no established criteria to guide validation have been established. Quality products need to be developed, and these must have defined performance parameters so that the users can apply the information generated effectively.

An event may occur where methods will be required that have not yet been validated by rigorously defined QA practices. For public security, it will be imperative that investigators make use of these techniques. If the results of one of these less-validated procedures are used for other than investigative leads, then a preliminary validation assessment should be carried out. One could convene a panel of experts, with proper security clearances, to assess the utility of the rapidly developed method and to define the limits of interpretation and conclusions.[35] Such an approach has been employed in the field of human DNA forensics for victim identifications in the collapse of the twin towers of the World Trade Center in New York caused by the terrorist acts of September 11, 2001.

Identifying pathogens to focus research efforts is a requisite. Many of the original top-priority biothreat agents were based on agents researched and weaponized in former state-sponsored bioweapons programs. Although, historically, food pathogens such as *Shigella dysenteriae*, *Salmonella typhimurium*, and *E. coli O157:H7* are readily accessible in nature and have been used,[10] little attention has been focused on agriculture, and plant pathogens in particular have been underrepresented in microbial forensic efforts. Although hundreds of potential pathogens and toxins exist, not all of them present the same threat as weapons. While one organism may have a high virulence, it may be very difficult to culture, disperse, or handle. Conversely, an organism with limited destructive potential may be readily available and easy to disseminate. To focus research efforts, it is therefore of crucial importance to outline a set of criteria with which to evaluate the potential threats posed by biological agents. Potential criteria include availability, ease of culture, survivability in the environment, ease of dispersion, infectivity, morbidity, impact on target population, immunity of target, therapeutic control measures, ability to be transmitted subject to subject, and ability to be bioengineered to increase virulence or to make the agent resistant to vaccines or therapeutics. The SWGMGF is considering all of these criteria and has added virulence, the knowledge base forming the foundation upon which biological marker development relies (i.e., state-of-the-art), the likelihood of use of a particular agent, and the impact of the use of the agent on social stability.

Once an organism is identified by whatever means is available, the significance of that finding needs to be conveyed appropriately. One should not interpret evidence beyond the limits of the assay. While inferences can be made without a complete understanding of the genetic diversity of an organism, more data will provide increased resolution toward the goal of attribution. This requires more population genetics and statistics studies. To date little has been done in the context of forensic attribution.

Information is the key to thwart terrorism and may be our best defense. Databases are required to have ready access to information. For a database to be effective, it must contain the appropriate data entry items, and the data must be retrievable. With proper planning, meaningful and useful databases can be developed. The database(s) criteria need to be defined, and SWGMGF is taking on this function as well. Criteria fields for information databases under consideration are: multi-agency threat lists (virus, bacteria, fungi, protozoa, insects, toxins), recognizing synonyms, strain data and virulence, vaccine strains, DNA sequences (whole genomic and partial), bioengineering events, biomarkers, organism sources (laboratory, geographic, and natural), national experts and contact information, laboratory facilities, assays, scientific presentations, literature references and full text, research grants, characteristics (e.g., microscopic morphology, colony morphology), antibiotic resistance, natural transmission, viability (e.g., aerosol, liquid), materials found with microorganisms (e.g., manufacturing process residue/metabolites, culture media, additives for processing, additives for stability), methods of manufacture, handling, packaging, shipping, sources for materials, environmental incidentals (e.g., pollen, fungi), historical outbreaks, known instances of threats or usage, hoaxes, epidemiology, strain and disease history, disease symptoms, dissemination strategies, dispersal strategies, methods for investigation prior to an event, and links to other databases. The Bioforensics Demonstration and Application Project supported by the Department of Energy (DOE) is developing a National Bioforensic Information Encyclopedia and National Bioforensic Evidence Database through the efforts of Lawrence Livermore National Laboratory, Los Alamos National Laboratory, and Northern Arizona University. This effort is coordinating with the SWGMGF to set criteria and functionalities. Obviously, the diversity of data entry items in such databases makes it essential that such databases include heteroformatic components and a relational structure (with appropriate links), in which all of the information on any specific record may not physically reside together. Therefore, one task of the database subgroup of the SWGMGF is also to define and suggest research and development needs for bioinformatic tools of suitable search engines for data retrieval in user-friendly platforms.

EPIDEMIOLOGY (A FIRST STEP IN A MICROBIAL FORENSIC ANALYSIS)

Successful use of a bioweapon may not become apparent immediately. Since it may take days to weeks after exposure for an individual(s) to develop symptoms, the first evidence of an event may well be cases in hospital emergency rooms. However, emergency room personnel may not be the first to detect a problem. Other indicators could be, for example, pharmacists distributing more antibiotics than usual, 911 operators experiencing an increase in health-related distress calls, and funeral homes having increased business.[36] Thus, better public health (and agriculture) surveillance is needed. Disease outbreaks, especially suspicious ones, should be evaluated as a potential bioterrorist attack. Epidemiologic tools can assist in differentiating between intentional and natural outbreaks.

An important step in achieving the goals of microbial forensics is the integration of different disciplines. It is important to recognize the role of epidemiology in forensics. A biological crime may be mistaken as a natural event. For events that have already occurred, epidemiologists will often be the ones to gather the data. The epidemiologist addresses the cause of the disease and treatment of the victim/patient. The evidence and data obtained at the initiation of the event also will be crucial for the forensic scientist to continue the investigation further for attribution.

The basic epidemiologic approach in the evaluation of a potential bioterrorist attack or biocrime is essentially the same as that of a natural outbreak.[36] The presence of some pathogens will automatically indicate an attack, such as smallpox.[35] A naturally occurring disease that is found outside its typical environment may raise suspicion and initiate a forensic investigation, such as a man in Florida contracting pulmonary anthrax with no known contact with livestock and a couple in New York developing plague (although the latter was not the result of a bioattack). Laboratory and clinical data are used to confirm that a disease outbreak has occurred. Then to trace back to the source, aspects of the health surveillance system are used, such as number of emergency visits, laboratory data, pharmacy use, work and/or school absenteeism, or any other data that correlate with an increase in infectious disease. It is obvious that an integrated surveillance system must be established, as it is essential for detection of emerging or reemerging disease. Microbial forensics requires established computerized surveillance networks to track infectious disease outbreaks in real time.[35] While some systems do exist, better connectivity is needed. Better integration with the CDC's Lab Response Network and PulseNet (for food pathogens) is essential. In the same light, interaction with the veterinarian, Animal and Plant Health Inspection Service (APHIS), and

Agricultural Research Service (ARS) for agricultural targets should be integrated with law enforcement.

MICROBIAL FORENSIC TOOLS

Epidemiology and forensics have similar roles when investigating biocrimes. However, after the determination of the cause, doctors treat the patient. Forensic scientists proceed further with analyses. There are additional stringencies placed on the forensic investigation, which include maintenance of chain of custody of the evidence and more detailed analyses of the pathogen that was used as a weapon. If possible, the analyses may determine the strain or substrain of the organism and other features that may be unimportant to the public health sector.

While protocols exist for evidence collection and preservation, challenging samples will be encountered. Samples that include soil, mud, swamp water, large items that may not fit into biosafety containment hoods, and other types of materials must be considered. The BFAC and its partners must consider various scenarios and develop processes to handle a wide variety of samples and sample types, at times under the constraints of biocontainment.

Clearly DNA typing methodologies will figure prominently in the attribution of a pathogenic weapon. First, the DNA must be successfully extracted from a sample. Extraction procedures that are used for pristine samples may not be adequate for more challenging environmental samples. Efficiency of recovery may be low, and polymerase chain reaction (PCR) inhibitors may not be removed effectively. The BFAC and its partners will have to promote the development of better DNA extraction methods. Once DNA has been recovered, molecular biology tools and biomarkers exist to assist analyses; these include sequencing, microarray analyses, pathogenicity array analyses, single-nucleotide polymorphism (SNP) characterizations, 16S rRNA sequencing, variable number tandem repeat analysis, and antibiotic resistance gene characterizations. This list of analyses is impressive. However, more research is needed regarding the selection of the genomic region of pathogen agents to determine the severity of pathogenicity. Further, still more assays need to be developed for better resolution of strains and substrains of pathogens. Detecting bioengineered organisms is also extremely important. Inserting a toxin gene into a normally nonpathogenic organism is feasible. Not only can previously harmless organisms be endowed with virulence by genetic engineering, but certain virus vectors can also serve as carriers of harmful genes that could be inserted into a host's own cellular DNA. Analysis of the genome of an organism may be able to differentiate a bioattack from a natural outbreak. Toxins may be characterized by immunoassays, biofunctional assays, peptide or protein-based assays, or by mass spectrometry. In addition, trace levels of the DNA from the organism that produced the toxin may be present, and DNA-based assays could aid in attributing the source of the toxin.

As stated above, forensic investigations use all types of physical evidence to attempt to obtain information on the organism, the persons involved, the places involved, the processes used to develop or disperse the weapon, the instrumentation used, and/or the time of the criminal act. Other materials and methodologies may assist in an investigation of a bioterrorism event, and at times may be more informative than DNA-based tests. Traditional physical evidence collected at a crime scene will also be evaluated. In addition, a recent report by the American Academy of Microbiology[34] described a number of non-DNA-based approaches to consider for the microbial forensic panoply of assays. These include:

1. The physical attributes, such as morphology and microstructure, acquired by the microorganism(s) during preparation for weaponization may be distinctive and provide information on attribution;
2. Isotope analyses can be used to approximate the age and source of the microorganisms;
3. Traditional physiologic methods (e.g., fatty acid composition, phage typing, serotyping) may provide information to further identify the microorganisms;
4. Remnants of growth media and media components adhering to the microorganisms may provide clues to the source of the weapons or methods used to prepare the material;
5. Stabilizers and additives in the preparation of a sample may be signatures that can provide leads about the perpetrator;
6. Incidental biocontaminants, such as environmental pollen and fungi, may give clues to the location and time of year the sample was prepared;
7. Bacterial endemism is the existence of unique strains of bacteria that may exist in only one location or rarely in other locations. Identification of such incidental bacteria coexisting in a weaponized sample may be used to geolocate the sample source or place of preparation;
8. Monitoring changes in the immunological response of a host to a pathogen, such as temporal IgG and IgM responses to epitopes of the microorganism, may indicate whether or not an individual has had long-term or recent exposure to the microorganism or toxin;
9. Immunoassays are often used as rapid detection methods and are especially applicable to field-deployable assays. Antibodies may also be specific to particular protein signatures that could further resolve the identity of a material.

Analysis of all pertinent physical evidence collected at a crime scene will assist in attribution. It is of tantamount importance to consider other tools besides solely DNA-based assays. Since microorganisms often propagate clonally, it may be difficult to differentiate some samples based on genetics, especially slowly evolving species. These non-DNA-based assays and others are part of the suite of potential approaches that a microbial forensic scientist may be able to use for attribution today and in the future.

BIOINFORMATICS AND GENETICS

Genetic markers have been used for more than a decade in forensic identity testing to differentiate humans, animals (such as dogs, horses, and cows), and plants. For those diploid, sexually reproducing organisms, sufficient genetic variation exists to distinguish unrelated and related individuals (excluding identical twins). Thus, in a forensic identity test, the DNA in the sample serves as a signature for the organism. However, with microorganisms the DNA sequence may be the same or similar in multiple isolates.

Some microorganisms may carry forensically significant markers that may distinguish some isolates from others, such as some of the highly diverse enteric pathogens;[37,38] others such as *Bacillus anthracis* may not be readily distinguishable genetically among some strains of the species. Although such cases have a substantial difference in diversity, phylogenetics may be employed to infer relationships based on evolutionary characteristics. Obviously, the more derived characters organisms share, the more closely related they are. Thus, an understanding of genetic variation within a species and, if necessary, its near neighbors will improve capabilities to resolve attribution of isolates or samples. Major efforts in genome sequencing are underway, and the costs are decreasing substantially. So as long as resources are adequate and focus maintained, gathering the genetic information on priority bacteria and viruses should be possible. However, the genomes of fungi are substantially larger than bacteria, and more efforts will be needed to acquire the basic genetic information on significant fungal pathogens.

Data acquisition, manipulation, analysis, and presentation will figure prominently in microbial forensics. Bioinformatics, including comparative and functional genomics, is an important cornerstone in understanding whether an event is a natural or unnatural occurrence. Genetically engineered organisms may be able to be identified. The understanding of host-pathogen interactions will enable the scientist to evaluate the potential impact of a weaponized agent and the function of pathogenic genes and gene families. Computer models may be used to predict biology and improve assay design even if the use of the organism is restricted. Expression arrays will help identify genes and functions. Rational design of assays, instead of relying on only anonymous sequence data, will enhance detection and identification capabilities. Since the basic objectives of gathering such genomic data are to determine the extent of pathogenicity of the agent and subsequently make attempts towards source attribution, the annotation of relevant microbial genomic databases is extremely important. This area of bioinformatic research is still preliminary. However, model systems are emerging through annotation of the human transcriptome databases that the microbial forensic community may find useful.

FORENSIC CASE EXAMPLES

To better plan for forensic investigation strategies of an event, capitalizing on past experiences can shed light on current practices and the direction of future research. Two illustrative examples, neither of them bioterrorism acts, are the Sverdlovsk anthrax incident and the case of a dentist with AIDS who may have infected patients with HIV. Both required epidemiological investigation and molecular biology to resolve.

The first case explores the sources of anthrax spores from an anthrax epidemic. In April 1979 in Sverdlovsk, Russia, a loud explosion occurred at a military compound. Several days later, residents downwind from this compound developed high fever and had difficulty breathing. The persons affected lived or worked downwind and within a narrow zone, approximately 4 km south and east of the military facility. At least 200 patients died. Medical doctors identified the outbreak as pulmonary anthrax. Government officials, in contrast, declared that the outbreak was caused by consumption of meat from a cow suffering from the disease. Thus, there was a discrepancy regarding the cause of the anthrax epidemic; either it was due to consumption of tainted meat or it was the result of a release of aerosolized anthrax from a bioweapons factory (a Biological Weapons Convention treaty violation).

Patient fatalities did not display the symptoms of gastric or skin anthrax, which would likely have been the result of handling or eating contaminated beef. Investigations also revealed that sheep and cattle in six different villages as far as 50 km southeast of the facility also died of anthrax. Autopsies of human victims revealed severe pulmonary edema and toxemia. DNA analysis suggested that the patients were exposed to several anthrax strains simultaneously, which is atypical for a natural infection in a cow. The available evidence supported that the cause of the disease was anthrax from an accidental release from the compound, a Russian bioweapons facility.[39,40] The release of the anthrax spores apparently occurred only once and may have taken no more than a few minutes.

The second case describes attempts to determine the cause and source of HIV in patients that implicated a Florida dentist with AIDS. A number of people from Florida became HIV-infected in the late 1980s. Their histories did not indicate lifestyles or practices that would put them at high risk for exposure to HIV.[41] An epidemiologic investigation indicated that the HIV transmission occurred during invasive dental care from a dentist with AIDS. The dentist was first identified as HIV-positive in 1986.[42,43] The HIV-positive patients had invasive procedures performed by the dentist after he had been diagnosed with AIDS. The precise mode of HIV transmission was unknown; the only data implicating him as the source are that multiple patients who contracted HIV visited that same dentist.

To provide further insight into whether or not the dentist transmitted HIV to his patients, bioinformatics/phylogenetics were employed. DNA sequence data from the HIVs from lymphocytes from each of the patients, the dentist, a local control group, and an outgroup were aligned and compared. The basis of the phylogenetic comparison is that sequences that are more similar are more closely related and share a more recent common ancestry than the sequences from other lineages.[43,44] The analysis showed that the HIV nucleotide sequences from a number of the patients were closely related (although not exactly the same) to those from the dentist and were distinct from viruses obtained from control patients living in the same geographic area as the dental practice. The data strongly support (but not necessarily conclusively) that the patient HIV is closely related to the HIV from the dentist contracted the virus from the dentist.[45]

For other microorganisms that do not mutate as rapidly, more similar sequences may be required for an association. Further, a lower evolutionary rate of divergence suggests that a larger segment of the genome of many microbes must be sequenced (or at least typed for the informative sites). Thus, it is important to increase efforts in genetics studies to gain a better understanding of the degree of variation among pathogenic agents so that circumstantial evidence for attribution can be strengthened.

EDUCATION

One of the best ways to combat a bioterrorism event (or any disease outbreak) is by having an involved and educated public. Surveillance by those who recognize a disease (particularly a rarely occurring one) can help reduce the effects of the disease. Furthermore, proper understanding of and response to an event can reduce public fear, minimize economic impact, and halt the paralyzing effect that some terrorists attempt to inflict. However, the public may not be educated sufficiently to serve in this capacity. The following example of the state-of-understanding of a potential health risk and bioterrorism act demonstrates the need for better education tools and strategies.

The risk of a smallpox attack is unlikely. The known declared supplies of the variola virus supposedly reside only in secure facilities at the State Research Center of Virology and Biotechnology (called VECTOR) in Koltsovo, Russia and the U.S. (at the CDC).[46] However, it is possible that rogue states have isolates of the smallpox virus. Thus, the potential for a smallpox attack has led to a debate on what precautions should be taken nationally. To be effective, the public should be knowledgeable about smallpox and the risk of prophylaxis. Lack of knowledge will render the public less capable to identify a victim of smallpox (if one is ever encountered), to appreciate the degree and intensity of transmission, and to make judicious decisions on whether to receive the vaccine.

Blendon et al.[47] conducted a survey of randomly selected adults (n = 1006) to evaluate their knowledge of and beliefs about smallpox virus and the vaccine and their possible reactions to an attack. The majority of respondents had misunderstandings or wrong information about smallpox and smallpox vaccination. Some beliefs were that (1) there had been a case of smallpox in the last five years (the last in the world was in 1977); (2) there is an effective treatment for smallpox once contracted (although none exists); (3) vaccination within two to three days after exposure does not provide protection (although it does, assuming the virus has not been engineered to evade the immune response of the host);[48] and (4) there are not enough doses of vaccine for the public (although there are).[49]

While not a direct action of a microbial forensics program, the above described state of affairs of the public's understanding points to a need for better education. Smallpox has been presented substantially in the news, yet the public knows little about it. Increased preparedness and planning will enable early recognition of the disease, effective vaccination strategies, and proper isolation procedures to mitigate the spread of the disease.[45] Early detection will stem the spread of disease and allow better gathering of forensic evidence. Furthermore, there is no reason to believe that the public's (including public leaders) understanding of bioterrorism risks, prevention, and recognition is any better for other potential pathogenic weapons and their impacts on society. Without education, early detection/warning of an event may not occur, and out-of-scale reactions will likely occur. No modeling can predict the thoughts and actions of terrorists or how the public and its leaders will respond during times of crises.[50] However, education to some degree can place issues and actions in proper perspective.

The microbial forensics field, in collaboration with, for example, the American Society of Microbiology, must develop educational tools so that materials will be available for lectures, courses, workshops, and general dissemination to the public. The SWGMGF has deemed education an important part of its mission.

CONCLUSION

Advances in technology will no doubt develop better treatments and therapies for many of the microbial diseases that affect humans, animals, and plants. However, these same technologies can be abused and used as weapons; thus challenges to national (and international) security will continue to arise. Because it may be difficult to detect and deter use of a bioweapon, vigilance is necessary. Science can take up the mantel and provide tools to detect and attribute the pathogen and those who may use bioweapons to create terror or to commit crimes. A multidimensional formulation of the microbial forensics discipline, as outlined in this chapter, should be capable of addressing these aims. By developing a robust microbial forensics field, supported by the BFAC, a developed knowledge base; integration of government, academic, and private sector institutions; and the standing working group, SWGMGF, we can be ready to meet the needs of society.

ACKNOWLEDGMENT

This is publication number 03–08 of the Laboratory Division of the Federal Bureau of Investigation with contributions from the Armed Forces Institute of Pathology. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

REFERENCES

1. Atlas RM. Bioterrorism: From threat to reality. *Ann Rev Microbiol* 2002;56:167–185.
2. Pearson GS. The threat of deliberate disease in the 21st century. Henry L Stimson Center Report, No. 24. Washington, DC: Stimson Center; 1998.
3. Christopher GW, Cieslak TJ, Pavlin JA, Eitzen EM. Biological warfare: A historical perspective. *JAMA* 1997;278:412–417.
4. Stockholm International Peace Research Institute (SPIRI). *The Rise of CB Weapons: The Problem of Chemical and Biological Warfare*. New York, NY: Humanities Press; 1971.
5. CDC website: www.cdc.gov/agent/agentlist.asp
6. Derbes VJ. De Mussis and the great plague of 1348: A forgotten episode of bacteriological war. *JAMA* 1966;196:59–62.
7. http://www.usamriid.army.mil/content/BioWarCourse/HISTORY/HISTORY.html
8. King PZ. Bacteriological warfare. *Chin Med J* 1943;61:259–263.

9. Broad WJ, Miller J. The threat of germ weapons is rising. Fear, too. *New York Times* 1998 Dec 27; sec. 4, col. 1, pg. 1.
10. Carus SW. *Bioterrorism and Biocrimes: The Illicit Use of Biological Agents Since 1900*. The Netherlands: Fredonia Books; 2000.
11. Keim P, Smith KL, Keys C, Takahashi H, Kurata T, Kaufmann A. Molecular investigation of the Aum Shinrikyo anthrax release in Kameido, Japan. *J Clin Microbiol* 2001;39:4566–4567.
12. Török TJ, Tauxe RV, Wise RP, Livengood JR, Sokolow R, Mauvais S, Birkness KA, Skeels MR, Horan JM, Foster LR. A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. *JAMA* 1997;278:389–395.
13. Kolavic SA, Kimura A, Simons SL, Slutsker L, Barth S, Haley CE. An outbreak of shigella dysenteriae type 2 among laboratory workers due to intentional food contamination. *JAMA* 1997;278:396–398.
14. OIE website: www.oie.int
15. National Research Council. *Making the Nation Safer. The Role of Science and Technology in Countering Terrorism*. Washington, DC: The National Academy Press; 2002.
16. Robertson AG, Robertson LJ. From asps to allegations: Biological warfare in history. *Mil Med* 1995;160:369–373.
17. National Research Council. *Countering Agricultural Bioterrorism*. Washington, DC: The National Academy Press; 2002.
18. Wilson TM, Logan-Henfrey L, Weller R, Kellman B. Agroterrorism, Biological Crimes, and Biological Warfare Targeting Animal Agriculture. In Brown C, Bolin C (eds). *Emerging Diseases of Animals*. Washington, DC: ASM Press; 2000:23–57.
19. Richard JL, Payne GA, Desjardins AE, Maragos C, Norred WP, Pestka JJ, Phillips TD, van Egmond HP, Vardon PJ, Whitaker TB, Wood G. *Mycotoxins. Risks in Plant, Animal, and Human Systems* Task Force Report, No. 139. Ames, Iowa: Council for Agricultural Science and Technology; 2002.
20. Yang XB, Dowler WM, Tschanz AT. A simulation model for assessing soybean rust epidemics. *J Phytopathol* 1991;133:187–200.
21. Yang XB, Royer MH, Tschanz AT, Tsai B. Analysis and quantification of soybean rust epidemics from seventy-three sequential planting experiments. *Phytopathology* 1990;80:1421–1427.
22. Whitby S, Rogers P. Anticrop biological warfare—Implications of the Iraqi and US programmes. *Defense Analysis* 1997;13:303–318.
23. Rogers P, Whitby S, Dando M. Biological warfare against crops. *Sci Amer* 1999;280(6):70–75.
24. Whitby SM. The potential use of plant pathogens against crops. *Microbes Infect* 2001;3(1):73–80.
25. Arnon SS, Schechter R, Inglesby TV, Henderson DA, Bartlett JG, Ascher MS, Eitzen E, Fine AD, Hayer J, Layton M, Lillibridge S, Osterholm MT, O'Toole T, Parker G, Perl TM, Russell PK, Swerdlow DL, Tonat K. Botulinum toxin as a biological weapon. In Henderson DA, Inglesby TV, O'Toole T (eds). *Bioterrorism–Guidelines for Medical and Public Health Management*. Chicago: American Medical Association Press; 2002:141–165.
26. Eitzen E, Pavlin J, Cieslak T, Christopher G, Culpepper R (eds). *Medical Management of Biological Casualties Handbook*, Third Edition. Fort Detrick, Frederick, Maryland: U.S. Army Medical Research Institute of Infectious Diseases; 1998.
27. Stone R. Peering into the shadows: Iraq's Bioweapons program. *Science* 2002;297:1110–1112.
28. Saferstein R (ed). *Criminalistics—An Introduction to Forensic Science*, Fifth Edition. New Jersey: Prentice Hall Education; 1995.
29. Possession, use and transfer of select agents and toxins. 42 CFR Part 73. Department of Health and Human Services; 2003.
30. Budowle B. TWGDAM guidelines for quality assurance program for DNA analysis. *Crime Lab Dig* 1995;22:21–50.

31. Budowle B, Smith JA, Moretti T, DiZinno J. *DNA Typing Protocols: Molecular Biology and Forensic Analysis*. Natick, MA: BioTechniques Books, BioForensic Sciences Series, Eaton Publishing; 2000.
32. CLIA ’88. Code of Federal Regulations. 57 CFR 7139, Washington, DC, U.S. Government Printing Office; 2001:883–999.
33. International Standards Organization (ISO)/International Electrotechnical Commission (IEC), ISO/IEC 17025, General requirements for the competence of testing and calibration laboratories. American National Standards Institute, New York, NY; 1999.
34. Budowle B, Schutzer SE, Einseln A, Kelley LC, Walsh AC, Smith JAL, Marrone BL, Robertson S, Campos J. Building microbial forensics as a response to bioterrorism. *Science* 2003;301: 1852–1853.
35. Keim P. *Microbial Forensics: A Scientific Assessment*. Washington, DC: American Academy of Microbiology; 2003.
36. Danzig R, Berkowsky PB. Why should we be concerned about biological warfare? *JAMA* 1997;285:431–432.
37. Brown EW, Kotewicz ML, Cebula TA. Detection of recombination among *Salmonella enterica* strains using the incongruence length difference test. *Mol Phylogen Evol* 2002;24:102–120.
38. Kotewicz ML, Brown EW, LeClerc JE, Cebula TA. Genomic variability among enteric pathogens: The case of the mutS-rpoS intergenic region. *Trends in Microbiology* 2003;11:2–6.
39. Marshall E. Sverdlovsk: Anthrax capital? *Science* 1988;240:383–385.
40. Meselson M, Guillemin V, Hugh-Jones M, Langmuir A, Popova I, Shelokov A, Yampolskaya O. The Sverdlovsk anthrax outbreak of 1979. *Science* 1994;266:1202–1208.
41. Ciesielski C, Marianos D, Ou CY, Dumbaugh R, Witte J, Berkelman R, Gooch B, Myers G, Luo CC, Schochetman G, et al. Transmission of human immunodeficiency virus in a dental practice. *Ann Intern Med* 1992;116:798–805.
42. CDC. Possible transmission of human immunodeficiency virus to a patient during an invasive dental procedure. *MMWR* 1990;39:489–493.
43. CDC. Update: Transmission of HIV infection during an invasive dental procedure—Florida. *MMWR* 1991;40:21–27.
44. Ou CY, Kwok S, Mitchell SW, Mack DH, Sninsky JJ, Krebs JW, Feorino P, Warfield D, Schochetman G, et al. DNA amplification for direct detection of HIV-1 in DNA of peripheral mononuclear cells. *Science* 1988;239:295–297.
45. Ou CY, Ciesislski CA, Myers G, Bandea CI, Luo CC, Korber BT, Mullins BT, Schochetman G, Berkelman RL, Economou AN, et al. Molecular epidemiology of HIV transmission in a dental practice. *Science* 1992;256:1165–1171.
46. Bozette SA, Boer R, Bhatnagar V, Brower JL, Keeler EB, Morton SC, Stoto MA. A model for a smallpox-vaccination policy. *New Engl J Med* 2003;348:416–425.
47. Blendon BJ, DesRoches CM, Benson JM, Herrmann MJ, Taylor-Clark K, Weldon KJ. The public and the smallpox threat. *New Engl J Med* 2003;348:426–432.
48. Mack T. A different view of smallpox and vaccination. *New Engl J Med* 2003;348:460–463.
49. Smallpox fact sheet, smallpox overview. Atlanta: Centers for Disease Control and Prevention; 2002 at http://www.cdc.gov/smallpox.
50. Schraeder TL, Campion EW. Smallpox vaccination—The call to arms. *New Engl J Med* 2003;348:381–382.

CHAPTER 2

Infectious Diseases: Not Just a Health Matter Anymore

ROGER G. BREEZE
Centaur Science Group, Washington, DC

INTRODUCTION

Barely 15 years ago, only a handful of scientists in the military, medical, and veterinary fields had more than a passing interest in infections such as smallpox, anthrax, plague, and tularemia. Even fewer in the agricultural sphere worked on foreign pathogens that posed threats to U.S. livestock populations or our major crop commodities. Three federal government departments whose missions encompassed preparedness for uncommon infections—Agriculture, Defense, and Health and Human Services (HHS)—supported these efforts, and their combined budgets for this effort were small. But in 2005 talk of these infections is on many lips, the list of federal and state government agencies with active programs has grown significantly, and the funds available have mushroomed.

The purpose of this chapter is to explain how this came about and why the intelligence community and the Departments of Homeland Security, Justice, and State have now joined the traditional agencies—Agriculture, Defense, and HHS—as significant players in studies of infectious diseases of humans, animals, and plants. Subsequent chapters in the book explain in more detail the specific needs of these new entrants to the field.

This is not the place to recount the history of the development and use of biological weapons, which has been fully detailed elsewhere.[1,2] The seminal events for our purposes occurred over the past 15 years; they are the realization that molecular techniques could transform the properties of microbial pathogens; discovery of the vast covert offensive biological weapons programs of the former Soviet Union; proliferation of biological weapons technologies to smaller countries that also harbored or supported organized international terrorist groups; attacks in Japan with biological and chemical weapons

Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

mounted by Aum Shinrikyo, a religious cult with access to skilled scientists and abundant finance; attacks in the U.S. by domestic groups or individuals employing biological agents, such as the followers of Bagwhan Rajneesh in The Dalles, Oregon, and conventional explosives, as used by Timothy McVeigh in Oklahoma City, Oklahoma; and finally the al Qaeda attacks of September 11, 2001 and the anthrax letter campaign of October 2001 for which the purpose and perpetrator(s) remain unknown at this date. Today, we find ourselves locked in what appears to be a lengthy global struggle against a number of foreign terrorist groups at the same time that we confront individuals and groups domestically who are prepared to employ biological agents to advance their goals.

The common thread is that nation states, organized groups, and individuals can now see that for their own purposes it is possible to employ a group of readily available viral, bacterial, and fungal pathogens and their toxins—either in their native form or after genetic manipulation–to cause very serious consequences for humans, livestock, or plants. Questions of who would do this, why, and how have many answers—too many and too diverse to relate here.[1,2] It is only important to realize that there are three concepts: biological warfare waged by nation states as an act of war; biological terrorism employed by terrorist groups that may or may not be supported by a nation state; and biological crimes (biocrimes) performed by individuals or small groups for nonwarfare or nonpolitical purposes.

Biological warfare—the use of disease-spreading microorganisms, toxins, and pests against enemy armed forces or civilians—as historically intended or conducted by nation states for military purposes, generally focused on infectious but noncontagious agents. These give rise to incapacitating or fatal illness in the human population exposed to the weapon, but pose no risk of rebounding on the attacker. Anthrax, tularemia, Brucella, and encephalitis viruses are the classic examples. The former Soviet Union's extensive development of smallpox biological weapons was an exception to this principle. Biological weapons targeting livestock or crops—such as the anthrax cattle feed cakes produced by Great Britain during World War II—were intended to deny the adversary's population the food supplies essential to continuing hostilities. We may still face the threat of biological weapons wielded by a nation state as an act of declared or undeclared war. But our military, political, and other means to counter such an event appear far greater than those we might employ against the much more dangerous yet amorphous threats of international terrorism.

The word *terror*, intense fear, is derived from the Latin verb *terrere*, meaning to frighten. There are many definitions of "bioterrorism." The one used here is: "the threat or use of biological agents by individuals or groups motivated by political, religious, ecological, or other ideological objectives."[1] Aum

Shinrikyo provides an example; another may be the 2001 anthrax mail attacks. These events are different from the deliberate misuse of biological agents in biocrimes—as described in Chapters 4 and 18.

INFECTIOUS DISEASE: THE CLASSIC HEALTH VIEWPOINT

Biological agents developed as weapons pose different threats than the naturally occurring disease states traditionally associated with these infections in humans, animals, or plants. These differences become clear by a comparison of how public health or agricultural disease professionals and those engaged in creating offensive biological weapons programs see the infectious disease process.

Infectious agents, such as viruses, are composed of a delivery system—the outer coat of the virus—and a payload—the genetic material inside the virus. Their target is the host cell. This is a classic weapons scenario that those engaged in offensive biological weapons development know and exploit to the maximum. As Peter and Jane Medawar so aptly noted, "a virus is just a piece of bad news wrapped in protein."

The interaction between an infectious agent and a susceptible host is quite familiar, perhaps too familiar, to the medical professions and those charged with protecting animal and plant health. It is encapsulated in Figure 2.1: a microorganism is either virulent or not, and its contact with a host that is susceptible or not governs whether illness and death follow. The health professions in their entirety cannot help but look at this equation from one perspective—that the outcome be health and that the host regain full function, life, and happiness—because our education and ethical systems are built on these values.

Because of this perspective, health professions orient their research programs in infectious diseases along familiar lines: they dissect the infectious agent to understand those parts that might be used in a vaccine, and they examine the course of the infection in the host to find where the virus can be detected for diagnosis or where it persists in chronic infections. Their goals are simple: to identify the properties of the virus and of the host that can be of value for the purposes of diagnosis, treatment, and prevention.

Smallpox is by far the most important biological weapons threat because the disease is highly contagious, spreads easily by contact or aerosol from one person to another, and most probably could not today be limited to one country or continent after deliberate release. *Variola* virus, the cause of smallpox, can also be readily altered by genetic engineering techniques to enhance

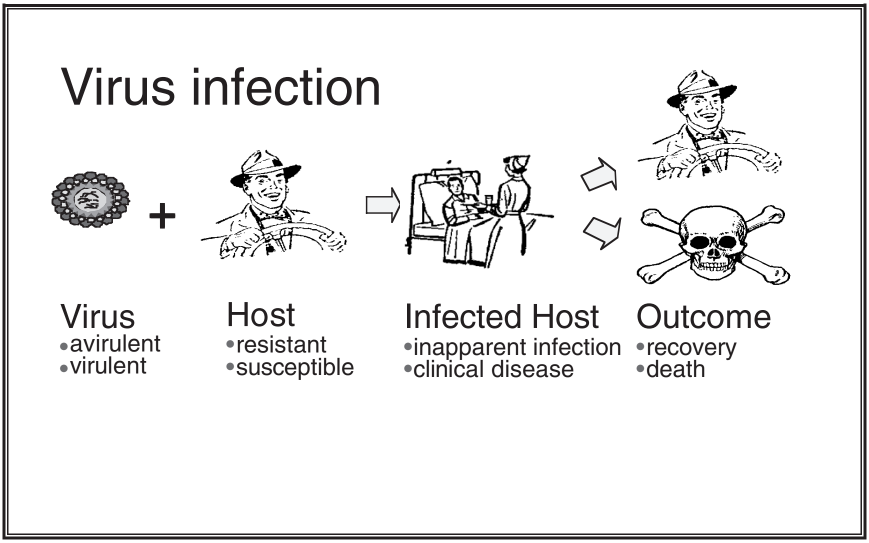


FIGURE 2.1 Consequences of the encounter between an infectious agent and a host. If the infectious agent is not virulent or the host is not susceptible, there is no infection and no disease.

its virulence. The day is not far off when this pathogen can be synthesized *de novo*. An attack with the smallpox virus on the United States would threaten the entire world, could result in tens of millions of deaths, and could paralyze the economies of industrialized nations. Given these possibilities, health professionals have difficulty even comprehending that someone would deliberately release such a devastating disease. When forced to confront the threat of smallpox as a biological weapon, most health professionals would stress the need for adequate public health preparedness and see the most urgent questions to be answered as these:

- Do physicians have the skills and knowledge to recognize smallpox?
- Can the disease be diagnosed quickly and accurately?
- Do we have enough smallpox vaccine for the civilian population?
- Are first-responders vaccinated?
- Can we make a safer vaccine for immunocompromised individuals?
- Can we treat adverse effects of vaccination?
- Do we have policies and procedures in place for preventive and emergency vaccination?

Solutions to all these very important public health questions will certainly be critical if we should ever be attacked with smallpox, and the prudent first step as a nation must be to strengthen the public health system to answer each question positively. However, these are not the only important questions for a biological weapons defense program.

INFECTIOUS DISEASE: THE OFFENSIVE BIOLOGICAL WEAPONS VIEWPOINT

Scientists engaged in offensive biological weapons development see pathogens as weapons, and they look at the same equation for the interaction between a pathogen and a host shown in Figure 2.1. However, they do this from a very different perspective and with very different outcomes in mind from those involved in health care and delivery. For example, developers of offensive biological weapons want to understand the properties of the weapon's delivery system that allow it to spread by aerosol and survive in unfavorable environments, and those attributes of the payload that confer ability to evade host defenses and cause injury and death. They also seek to understand and undermine the properties of the target that confer resistance and susceptibility. Their goal is to enhance the weapon's capacity to cause injury in the host, and their desired outcome is death or disability. Understanding how a virus enters a susceptible target and how the virus outwits or evades the defenses of the host are critical factors in biological weapons design.

For those who want to make weapons, there are two desired outcomes: the native organism is honed to maximize death and injury, and advanced weapons based on the native pathogen are created by genetic engineering technologies. Advanced weapons have new properties, such as additional foreign genes that change the native biological properties to alter the species or organ system target, to overcome vaccination, to obscure diagnosis, to enhance transmission, or to add completely unexpected physiological effects. Advanced biological weapons pose new threats of technological surprise for the opponent. Wheelis[3] outlines the issues in more depth.

The questions those engaged in offensive smallpox biological weapons development might pose include:

- Can smallpox virus be engineered to cause disease in people who have been vaccinated?
- Can another pox virus be engineered that has the same effects as smallpox and evades vaccination?
- Can this pox virus be made highly infectious by aerosol?
- Can the smallpox virus be altered to have additional unsuspected effects that would result in more disability and death?
- Can the recognition of smallpox by physicians or its detection and diagnosis be confused or delayed so infection could spread?

Were any of these questions to be answered affirmatively and such weapons released, our traditional public health defenses would prove insufficient. Given the past and present existence of covert foreign offensive biological weapons programs, the inexorable march of science and technology that could put these dangerous pathogens, knowledge, and skills into more and more hands, and the apparently growing ranks of non-state entities prepared to use such weapons, a comprehensive program to defend the U.S. must counter offensive biological weapons challenges and ensure public health preparedness. These are the factors that have brought new government players to the field.

INFECTIOUS DISEASE: NEW PLAYERS AND NEW THREATS

DEPARTMENT OF DEFENSE AND THE INTELLIGENCE COMMUNITY

These agencies have always had the prime roles in defense against traditional biological warfare and have vigorously responded to changing threats over the past 15 years.

Defense focused first on real-time detection of biological weapons agents in air, water, and soil on the battlefield and translating detection into immediate effective protection of the Warfighter. Those biological agents long considered potential weapons were the first priorities. As the nature of the threat to our armed forces has become more diverse—all those persons on military bases at home and abroad, critical ports through which troops and material must load and unload—so Defense programs have expanded to provide a broader range of detection and countermeasures against a broader array of pathogens that might be encountered on and off the traditional battlefield.

The roles of intelligence community members will not be discussed in detail here. Suffice it to say that entities such as the Central Intelligence Agency are charged with using the equation shown in Figure 2.1 to discover those posing threats to us abroad, the nature of these threats, and the potential targets and means of delivery. Biology is just one of the technologies deployed for these purposes. The reader will appreciate the singular challenge posed in trying to identify a few people engaged in threatening activities among a global ocean of legal activities employing almost exactly the same types of persons, equipment, and facilities in academia, industry, and government. And even when suspects are identified it may not be possible to gain entry to the facilities they use, and thus remote means must be employed to discover what is going on inside. The kinds of biological questions the Agency might face could include:

- What pathogen is in the facility?
- Can that pathogen be reliably recovered from and identified in nonclinical materials from the suspect facility?
- Is reliable recovery possible after materials have been treated in ways adverse to pathogen survival?
- How few copies of the pathogen can be there for detection and how much of the whole organism is essential?
- Where did the pathogen come from?
- Has the organism been genetically altered and, if so, what unexpected properties have been added?
- Is there a foreign gene(s) present in the pathogen?
- What does the gene(s) do?
- Where did the gene(s) come from?
- How did the gene(s) get there?
- Is there any other evidence from the pathogen, the genome, or the delivery vehicle that would indicate laboratory origin or deliberate manipulation?

DEPARTMENT OF JUSTICE AND THE FEDERAL BUREAU OF INVESTIGATION

Deliberate attacks with biological weapons on the U.S. or U.S. interests at home and abroad would be either an act of war, if conducted by a nation state, or a crime. In either case, the Federal Bureau of Investigation (FBI) would be called upon to investigate. The FBI must thus look at Figure 2.1 with an eye to preparing for a wide variety of investigations: to determine what can be gleaned from studies of the biological agent, the victim and, perhaps, the perpetrator, to clear the innocent and to identify those responsible so that they may be successfully prosecuted.

This volume is largely devoted to the FBI's scientific needs for microbial forensics—the tools and technologies that will equip the Bureau to investigate biological crimes successfully. That such specialized and unmet needs exist at all will come as a surprise to the vast majority of readers, and the first purpose of this volume is to alert the wider scientific community to this deficiency. The chapters by Budowle and colleagues and by Harmon explain what the Bureau needs, how biological information will be used with other evidence to investigate and correctly attribute and prosecute criminal activity, and the standards that courts will demand for the scientific evidence. The rest of the volume starts to outline the field of microbial forensics through example and application of specific technologies. We intend that the volume as a whole should fulfill a second and greater purpose: to attract other scientists with novel ideas and skills to the field.

DEPARTMENT OF HOMELAND SECURITY

The FBI is charged with investigation of potentially criminal activity involving biological agents, during which the Bureau will use tools mostly developed by other agencies, of which the Department of Homeland Security (DHS) is the principal. The chapter by Budowle and colleagues explains the role of the DHS's Bioforensic Analysis Center as the core of a coordinated national activity.

DEPARTMENT OF STATE

The Department of State would take the lead in responding to biological attacks on U.S. citizens and interests overseas—other than those occurring as an act of war, which would fall under the aegis of the Defense Department. In addition, State is responsible for monitoring compliance with the Biological and Toxin Weapons Convention. To meet this mission, State must be capable of identifying suspicious disease outbreaks in humans, animals, and plants overseas that might be indicators of covert or illegal biological weapons activities undertaken by a nation state, organized group, or individuals. This is not a simple or rapid task, as illustrated by the Sverdlovsk anthrax outbreak described by Budowle and colleagues elsewhere in this volume. At a minimum, State must look at Figure 2.1 and be able to:

- Define how suspicious human, agricultural, and wildlife disease outbreaks differ from natural disease outbreaks in terms of surveillance, diagnostics, characterization, and attribution.
- Identify any special measures, qualifications, procedures, safeguards, technologies, or scientific approaches that must be used in investigating such outbreaks given the limited capabilities of established international human and agricultural health agencies.
- Field an effective investigative team that is backed by proven microbial and biological forensic techniques that can meet the scientific evidence requirements of international courts.

Currently the U.S. cannot provide these capabilities to the extent one would desire: the DHS's Bioforensic Analysis Center and the FBI will be leading U.S. capability- and capacity-building in the necessary areas. Nor are there international agencies that might fulfill these needs. International agencies monitoring livestock and plant health matters have very limited capabilities, and those for wildlife diseases are virtually nonexistent. Even the World Health Organization (WHO), probably the best resourced of any, maintains a healthy distance. A recent report[4] outlined that agency's anticipated role in any incident of intentional, malevolent use of biological or chemical agents. Key points included:

- WHO will focus on the possible public health consequences of such an incident, regardless of whether it is characterized as a deliberate act or a naturally occurring event.
- WHO advises strengthening public health and response activities with an emphasis on:
  - More effective national surveillance of outbreaks of illness
  - Better communications between responsible agencies and better coordination of their responses
  - Improved assessments of vulnerability and effective communication about risks to both professionals and the public
  - Preparation for handling the psychosocial consequences of the deliberate use of pathogens and chemicals to cause harm, and
  - Contingency plans for an enhanced response capability.

WHO also points out that: “Should the United Nations be called on to respond to a request to investigate, WHO could be asked to provide technical expertise or to make available its existing resources and mechanisms. Non-public health issues related to investigations of reports on possible use of chemical and bacteriological (biological) or toxin weapons, however, remain the responsibility of the United Nations.”

Clearly, the Bioforensic Analysis Center and FBI will be meeting an important need for the U.S. and setting a key precedent for the international community by fulfilling the research and development program outlined elsewhere by Budowle and his colleagues. This volume will be critical in informing the broader U.S. and international scientific communities of the opportunities for contributions in this important area.

Specifically, in terms of future research investment, attention should be drawn to a significant gap in our knowledge of disease agents in wildlife species, in which category I include biting insects and ticks. This is an area in which suspicious disease outbreaks should be expected, especially at habitation borders where humans are now encountering wildlife species—and pathogens—from which they were previously separated by geography.

In terms of known pathogens, this would include most of the top human threats (except smallpox), such as plague, tularemia, anthrax, botulism, hemorrhagic fever viruses, encephalitis viruses, Nipah and Hendra viruses, and other zoonotic pathogens, several of which are transmitted by mosquitoes, ticks, and other biting insects.

Wildlife species, including biting insects and ticks, are also the reservoir of many unknown infectious agents that may cause little or no illness in the wild species which have adapted to them by evolution. However, when these infectious agents enter a novel host, such as humans, they can provoke very serious and fatal disease. Examples include: Ebola virus (possibly in a chimpanzee reservoir), Hendra and Nipah viruses (maintained as inapparent infections in fruit-eating bats), and Hanta virus (maintained in mice). The interface between wildlife and humans is an area of great risk to human health normally, and is one of the most likely sites for future suspicious disease outbreak investigations. Incidents of fatal human diseases caused by presently unknown pathogens can be predicted in Africa and South America as people begin to live permanently on newly cleared land at the margins of virgin jungle where they are exposed to biting insects and other wild reservoirs of these infections. Paradoxically, however, in the past ten years, Hanta virus occurred in the American Southwest and Hendra and Nipah viruses appeared in Australia and Malaysia, respectively. The latter two viruses, representative of a new family of viruses whose other members we do not yet know, were carried by fruit-eating bats that had no direct contact with mammals: the infection was transmitted by feces from bats roosting over places where horses or pigs were housed.

It is worth noting that the unknown pathogens that cause serious diseases and death in wildlife species will likely not be detected or raise alarm by themselves. The trigger for attention will be when these pathogens spill over into humans (Hanta virus), or horses and then humans (Hendra virus), or domestic swine and then humans (Nipah virus). The tools and technologies to investigate disease in wildlife species will thus be developed primarily to investigate disease in humans or livestock caused by the same pathogens.

INFECTIOUS DISEASE: WHAT ARE THE THREATS?

There are many dangerous pathogens that could be used in deliberate attacks on humans, livestock, and plants. A question that is often asked is: “Which is the most important?” The answer is: “The next one to be employed against the U.S.” This is not a trite comment. Because several federal agencies are struggling to balance a finite amount of resources against a growing list of potential threats, we cannot approach this problem linearly.

There are various “lists” of priority pathogens, and these vary by the originating agency. Defense has a fairly short list of pathogens that can mostly be traced back to a historical state-supported offensive weapons program. HHS has a somewhat longer list that adds domestic terrorism precedents, such as ricin and Salmonella, and foreign viral diseases such as Rift Valley fever and

Nipah virus. Agriculture maintains a long list of animal and plant pathogens, but how much of a threat most of these pose for the U.S. awaits more critical analysis. For example, it is difficult to envisage how the U.S. could be terrorized by camel pox, a virus that infects only that species. In any event, in the case of foreign animal and plant diseases, it is only the ways we have chosen to respond to inadvertent disease introductions in the past—by mass slaughter of infected hosts and sweeping trade penalties—that enable terrorists to threaten us with them in the future.[5]

But however the lists are chopped and changed, it is clear that agencies such as the FBI and State are always likely to be faced with suspicious incidents that involve either unknown pathogens or pathogens for which the specific science base is not available. As a result, analytical techniques that can be developed and used "on the fly" will always be necessary.

INFECTIOUS DISEASE: BIOSAFETY AND BIOSECURITY

Those who work with infectious agents are familiar with biological safety standards developed by the Centers for Disease Control and Prevention to protect the health of laboratory staff by reducing or eliminating accidental exposure to human pathogens and to prevent the release of microorganisms from the laboratory into the environment.[6] These standards are based on facilities constructed and operated so as to contain the pathogens and on working practices in the laboratory that safely manage infectious agents. Many scientists thought these guidelines were stringent enough to prevent release of dangerous pathogens from the laboratory, but they did not consider the possibility of deliberate theft. During 1999 and 2000, informal attempts by government scientists to develop specific security standards could not achieve interagency consensus. The absence of precedents of theft and malicious use undermined attempts to make security a priority, even though in this regard many laboratories handling dangerous pathogens stacked up poorly against supermarkets and shopping malls.

The anthrax mail attacks in fall 2001 ushered in extensive security restrictions in all laboratories handling "select agents," a specific list of high-consequence pathogens defined by the Departments of HHS and Agriculture. These restrictions involved physical security standards, personnel surety programs, and tighter controls on inventory, shipment, and transport of pathogens. To the traditional base of "biological safety (Biosafety)," these new laws added another layer of "biological security (Biosecurity)," designed to protect biological agents against theft or sabotage. As a result, the biological research community has chafed under the unfamiliar burdens of operating in a security-conscious environment. These burdens are unlikely to lift soon given increased public awareness of biological weapons and the threat of terrorism, and a belated realization that diagnostic and research laboratories are potential sources of viable, virulent pathogens and their toxins. Nevertheless, there is no doubt that in the rush to do something about security, most relied far too heavily on guns, guards, and gates.

The biological community needs specific new tools to achieve a balance between adequately monitoring certain biological agents and toxins (far fewer than are on the lists today) and not obstructing legitimate research and development involving those agents and toxins. A good place to start would be a reappraisal of the biological agents subject to restriction: not all agents need the same level of scrutiny, and the ones requiring the tightest safeguards should be those that are most attractive to terrorists because they are the most devastating when used as weapons. The next step should be a proper Biosecurity Risk Assessment that starts with a definition of the facility's biological inventory and evaluates the consequences of its loss, a step that enables the pathogens to be prioritized based on those consequences. The identities of those who might seek to steal the assets, along with their motives and methods can then be assessed. Then the overall risk—probability and consequences—of these undesirable events can be evaluated.

Microbial forensics in support of law enforcement must become a critical new tool. Specifically, if we can move beyond simple statements about lists of microorganisms possessed by a particular facility to the stage where each isolate in the inventory is characterized and "fingerprinted" (bearing in mind the limitations of this analogy described elsewhere in this volume), we will be able to move quickly in an emergency to differentiate those who are definitely not involved from those who could be. For example, the DHS's Plum Island Animal Disease Center is the only facility in the U.S. legally allowed to possess certain high-consequence livestock pathogens, most of which are viruses. In the event of an outbreak of one of these diseases in the U.S., the first suspicion today would not be an inadvertent introduction but a terrorist act. The second thought would be that someone at Plum Island might be involved. At the very same time that the U.S. would need the services of its handful of experts, they could be under suspicion. The ability to "fingerprint" immediately the outbreak virus and compare this with the fingerprints of those viruses stored at Plum Island would resolve this conundrum. Clearly, for the vast majority of biological laboratories and pathogens, the issue is not so simple. However, we can apply the same principles to a whittled down list of high-consequence pathogens, given application of the technologies described in this volume. There is a long way to go, but the road ahead is clearer than ever before.

REFERENCES

1. Carus WS. Unlawful Acquisition and Use of Biological Agents. In Lederberg J (ed). *Biological Weapons: Limiting the Threat*. Cambridge: MIT Press; 1999:211–231.
2. Carus WS. *Bioterrorism and Biocrimes: The Illicit Use of Biological Agents in the 20th Century*. Center for Counterproliferation Research, National Defense University, Washington, DC, July 1999.
3. Wheelis M. Will the "New Biology" lead to new weapons? *Arms Control Today* 2004;34:6–13.
4. World Health Organization. Deliberate use of biological and chemical agents to cause harm: Public health response WHO 55th World Health Assembly, A55/20, provisional agenda item 13.15, April 16, 2002.
5. Breeze RG. Agroterrorism: betting far more than the farm. *Biosecurity and Bioterrorism* 2004;2: 251–264.
6. Centers for Disease Control and Prevention, National Institutes of Health, Centers for Disease Control. Biosafety in microbiological and biomedical laboratories. Washington: U.S. Dept. of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention; 1999.

CHAPTER 3

The Fundamentals of Human Virology

JEFFREY WILUSZ
Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado

This chapter provides an overview of basic virology along with a primer on the numerous viruses that cause significant human disease. Viruses are grouped in terms of their genomic organization in order to facilitate a discussion of both the molecular and pathogenic aspects of their biology. In addition to increasing our understanding of mechanisms of pathogenesis and rational drug design, the detailed molecular characterization of these pathogens has provided the potential to identify molecular signatures that can be used to trace outbreaks of viral disease.

INTRODUCTION

Viruses were established as agents that cause human disease at the beginning of the 20th century. Their small size (approximately 2- to 60-fold smaller than a standard Gram-positive *staphylococcus* or *streptococcus* bacterium) gave viruses the distinct property of being able to pass through the conventional filters of the day, allowing their identification as "filterable agents." Today viruses have been identified that affect every kingdom—animals, plants, and bacteria. These ubiquitous agents therefore have the potential to influence the entire biota of the planet. While this chapter will focus on human viruses, the general principles discussed can be applied to other virus families.

Viruses are molecular pathogens. They possess no metabolism of their own and can be thought of as molecular parasites. A conventional virus is made up of two or three major components. A nucleic acid genome, which can be DNA or RNA (single- or double-stranded, contiguous or segmented) contains all of the genetic information and is responsible for encoding all of the

Microbial Forensics
Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

virus-specific macromolecules of the pathogen. Due to the aggressive application of molecular biological techniques, the sequences of many viruses are known and fully annotated. This nucleic acid genome is packaged in a protein shell called a capsid. The capsid proteins generally self-assemble to form the shell, which takes on an icosahedral, helical, or complex symmetry depending on the virus family. The functions of this capsid include packaging of the genome, protecting the genome from environmental insults, and effectively delivering the nucleic acid to the inside of living cells. Most RNA viruses and some DNA viruses are surrounded by a lipid envelope that is derived from the host cell. The membranes of enveloped viruses all contain viral-specific glycoproteins that aid in viral tropism.

It is interesting to note that the conventional virus described above is not the simplest molecular pathogen. RNA-only agents called viroids exist that can kill plants without ever making a protein. In addition, protein-only pathogens called prions "replicate" in the absence of a nucleic acid genome by inducing a conformational change in other normal prp proteins. Prions cause degenerative neurological diseases called spongiform encephalopathies in humans. The bottom line is that it is truly a molecular jungle out there.

There are eight basic steps in a viral infection of a eukaryotic cell. The first is adsorption of the virion to the cell surface through an interaction of viral proteins with specific cellular receptors and, in some cases, coreceptor molecules. Viruses can specifically and tightly interact with proteins (e.g., HIV and CD4), carbohydrates (e.g., influenza virus and sialic acid), and lipids (e.g., B19 parvovirus and globoside). This virus-receptor interaction in large part determines the tropism of many viruses. Preventing this interaction is the goal of neutralizing antibodies that are generated by the immune system. The second step is viral penetration of the cell surface. This can be done by direct membrane fusion or endocytosis. The third step is uncoating of the viral genome inside the cell. In many cases, uncoating is due to pH changes that occur inside endocytotic vesicles that result in structural rearrangements of viral surface proteins. The fourth step is primary transcription/gene expression. Many viruses are designed to actively express their genomes immediately upon uncoating. This is accomplished via several strategies, including the packaging of viral polymerases, packaging or production of strong transcriptional transactivation proteins, and the simple fact that some viral genomes can serve directly as mRNAs. Step five is replication of the viral genome. Since viral-encoded proteins play a key role in this step for many viruses, this step has been a prime target for the development of antivirals. The next step is called "secondary transcription," or gene expression that occurs off of progeny genomes. This is an important step that viruses use to amplify their gene expression to obtain maximal yields of viral progeny. It should also be noted that many viruses make different sets of proteins at early and late times postinfection. Early proteins generally include replication and transcription factors as well as proteins that allow the virus to usurp cellular metabolism. Late protein production focuses on virion structural components. The seventh step is the packaging of progeny viral genomes. Viral capsids in most cases self-assemble, and the genomic nucleic acid must be properly inserted. The final step is the release of progeny virions from the cell. For enveloped viruses, this involves budding and the acquisition of a lipid bilayer membrane.

There are several ways to determine whether a cell is infected with a virus. First, many viral infections elicit cytopathic effects such as the rounding of cells, shrinkage, aggregation, lysis, or cell-cell fusion. These effects are due in large part to the expression of specific viral proteins or the shut-down of host cell macromolecular synthesis that occurs during the infection. Second, focal points of viral replication and assembly can sometimes be observed. These are referred to as inclusion bodies; a classic example is the detection of Negri bodies (cytoplasmic acidophilic inclusion bodies) in rabies virus-infected samples. Third, viral proteins expressed on the surface of an infected cell can sometimes bind many red blood cells, a phenomenon called "hemadsorption" that can be very striking when observed in a light microscope. This technique can be very useful if a viral infection shows very little cytopathic effect. The fourth way to directly detect viral-specific gene products is through the use of antibodies in western blots, immunofluorescence, or enzyme-linked immunosorbent assays (ELISAs). Finally, polymerase chain reaction (PCR) with virus-specific primers can be used to directly identify viral-infected tissue. The use of PCR-based assays as a powerful molecular epidemiologic tool is discussed further below.

The quantitation of virus in a sample can be approached using either physical- or biological-based assays. Physical methods such as particle counts using an electron microscope, hemaglutination assays, and ELISAs/(RIAs) can provide an approximation of the total number of viral particles in a sample, but do not address the fact that for many viruses, the particle-to-biologically active particle ratio is rather low. Therefore biological assays that include serial dilutions in conjunction with plaque assays, focus-forming assays, or determination of the infectious dose needed to kill 50% of a target laboratory animal test group are often considered the gold standard.

There are several key definitions and unique aspects of viral genetics that should be stressed. First, a field or street isolate of a virus is obtained directly from the natural host. Viruses that are passaged in cell culture often adapt to growth under *in vitro* conditions and may lose or gain some properties relative to the field isolate. Second, co-infection with two or more viruses that contain a segmented genome can result in a reassortment of segments in the progeny virions. This can result in dramatic changes in the biologic and/or antigenic properties of the new viruses. Third, while DNA viruses have a mutation rate of 10^{-8} to 10^{-11} per nucleotide incorporated, RNA viruses have a million-fold greater mutation rate (10^{-3} to 10^{-4}). This is due to the presence of error-prone polymerases that lack the ability to proofread. Fourth, phenotypic mixing or pseudotype formation can occur in an infection by two different viruses when progeny viruses are produced that contain the capsid of virus “A” surrounding the nucleic acid genome of virus “B.” In other words, phenotypic mixing is the mixing of nonnucleic acid components between two viruses. Finally, many viruses generate defective progeny viruses as a normal part of their life cycle. While this is one reason for the low particle-to-plaque-forming-unit ratios that are observed in viral preparations, the production of defective particles can have biological consequences as well. It is well established for several RNA viruses that defective particles are naturally produced that interfere with the replication of the wild-type virus. These defective interfering particles moderate the course of an infection and can support the establishment of persistent infections.

One of the best ways to control viral infections in a susceptible population is to prevent them from occurring through the judicious use of vaccines. Vaccines have reduced the incidence of measles, mumps, and rubella by >99.7% in the U.S., and have completely eliminated natural transmission of smallpox and poliovirus. The three main types of vaccines currently in use are killed (i.e., formalin-fixed viruses), live attenuated vaccine strains, and subunit vaccines made from recombinant proteins. Additional types of vaccines, including DNA-based vaccines and use of single vectors for broad vaccination, may well be on the horizon.

Historically, the availability of effective antiviral drugs has lagged behind that of other antimicrobials. However due to advances in molecular virology and rational drug design, the arsenal of antiviral compounds is growing quickly. The prime targets of available antivirals include uncoating (e.g., amantadine prevents intraviral pH changes in influenza virions), viral polymerases to prevent replication (e.g., AZT, ribavirin, and acyclovir), viral proteases to prevent maturation of virions, neuraminidases to prevent viral release from cells, and cellular enzymes involved in protein synthesis (interferons). The application of antibodies for the treatment or prevention of viral diseases is also becoming more common. In summary, the antiviral arena is definitely an evolving field.

The International Committee on the Taxonomy of Viruses has classified the major human viral pathogens into approximately 21 families. Since more viruses remain to be discovered and the relationships of individual viruses to human disease continue to expand, this number is likely an underestimate. Viral classifications are based on a variety of factors. First is morphology—the size of the particle, its shape, and the presence or absence of membranes. Other factors include the properties of the viral genome, viral genomic organization and expression patterns, antigenic consideration, and biological relationships such as host range, etc.

AN OVERVIEW OF THE DNA VIRUSES

Of the ~21 families of viruses that cause common human diseases, only six possess a DNA genome. Consistent with the location of cellular enzymes that act on DNA, all of these DNA viruses have a nuclear life cycle, with the exception of one. The Poxviruses remain cytoplasmic during infection, likely because they are too large to fit through the nuclear pore. Of the nuclear DNA viruses, only the herpesviruses are enveloped. Finally, epidemiological studies suggest that infection by many of the DNA viruses is very common in the human population. An overview of the six families of DNA viruses based on their molecular properties and life cycles is presented below.

Parvoviruses contain a small but unique 5–6-kb single-stranded DNA genome. The key human pathogenic parvovirus is called B19, which specifically infects red blood cell precursor cells via the P antigen as its receptor. This results in a transient, usually subclinical anemia that can be rather severe in hemolytic anemia patients (i.e., sickle cell disease) or the immune-compromised. Several weeks later, an immune-mediated rash (erythema infectiosum or fifth disease) or arthritis-like symptoms develop. Another parvovirus that is worth mentioning due to its utility in gene therapy is adeno-associated virus (AAV). As its name implies, the virus is dependent on co-infection with another DNA virus for efficient replication. Due to the facts that AAV causes no recognized human disease and integrates at a predominant location in the cellular genome, it is currently under development as a means of therapeutic gene delivery to human cells. In addition to therapeutic uses, the possibility clearly exists that AAV can be engineered as a biowarfare agent as well (i.e., potent toxin genes can be inserted into its genome and effectively delivered to human cells).

Adenoviruses were discovered about 50 years ago and generally cause respiratory and ocular infections. Two of the approximately 50 serotypes (types 40 and 41) are a major agent of gastrointestinal distress. Focal outbreaks of adenoviral-induced disease (e.g., pharyngoconjunctival fever associated with contaminated swimming pools or ponds) can occur in the U.S. The plethora of serotypes makes vaccination impractical given current vaccine technology. Gutted adenovirus vectors are also under intense development as a delivery system for gene therapy. As noted above for AAV, adenoviral vectors can also be used to potentially deliver detrimental genes as well as therapeutic ones.

There are eight major human herpesviruses, seven of which cause defined human disease. Individuals will maintain herpesvirus infections in a latent form following the initial acute infection. Reactivation can occur under a variety of circumstances. Herpes simplex viruses type I and II cause ulcerative lesions of the skin and remain latent in nerve ganglia. Cold sores are due to herpes simplex virus reactivation from the trigeminal ganglia. Varicella zoster virus causes chicken pox during primary infection and zoster or shingles upon reactivation. Human cytomegalovirus is a major cause of congenital infections that result in a variety of symptoms including retardation and hearing loss. Epstein Barr virus (EBV) targets B cells and causes infectious mononucleosis in young adults. In addition, EBV infection is associated with cancers in some populations (Burkitt's lymphoma in Africa and nasopharyngeal carcinoma in Asia). Human herpes virus type 6 causes the common childhood rash roseola (exanthem subitum). HHV8 or Kaposi's Sarcoma Herpes Virus is a sexually transmitted agent that causes kaposi lesions in the immunosuppressed. In addition to vaccines for VZV, herpesvirus infections can be controlled by a highly effective antiviral acyclovir and its derivatives. The drug specifically targets the viral thymidine kinase gene and prevents replication during active infections. The widespread nature of most herpesviruses makes suspicious outbreaks less likely. It should be noted, however, that due to the sexually transmitted nature of HHV8, kaposi sarcoma was one of the defining lesions on homosexual AIDS patients noted early on in the HIV epidemic.

Papovaviruses contain a unique circular double-stranded DNA genome. The family consists of two human polyoma viruses, BK and JC, which are rather ubiquitous agents that often cause inapparent primary infections. Reactivation of these agents by immunosuppression, however, is associated with cystitis for BK and a severe neurological disorder, progressive multifocal leukoencephalopathy, for JC. The broad array of wart-causing agents, the human papillomaviruses (HPV), are also members of this family. In addition to warts and condylomas, it is important to note that the vast majority of cervical dysplasias observed in PAP smears are due to HPV infection. Infections by certain HPV subtypes, i.e. HPV16 and HPV18, are strongly associated with the development of cervical cancers and other malignancies.

The Hepadnavirus family contains a one key member Hepatitis B virus (HBV). HBV is a unique agent that replicates using an RNA intermediate and reverse transcriptase. It is a leading cause of chronic hepatitis outside of the U.S. An effective recombinant subunit-based vaccine is currently available.

Poxviruses are exceptionally large enveloped virions with a unique complex viral capsid structure. Their entirely cytoplasmic life cycle is unique among the human DNA virus families. The key human pathogens in this group are variola (smallpox) and mulluscum contagiosum (which causes wart-like lesions). Smallpox was eradicated many years ago due to the effective application of the smallpox vaccine (vaccinia virus) and the fact that humans are the only natural reservoir of the agent. The risk of smallpox as a biowarfare agent has received much recent attention due to the declining or total lack of immunity currently seen in the population. Finally, the 2003 outbreak of monkeypox, a nonfatal rash disease, in the upper Midwest that was previously unseen in the western hemisphere serves as a reminder that other zoonotic viruses exist that can emerge in the human population.

AN OVERVIEW OF THE RNA VIRUSES

Fifteen families of the 21 virus families that cause major human diseases contain an RNA genome. Four useful generalities can be applied to virus families that contain RNA genomes. All of the RNA virus families maintain an exclusively cytoplasmic life cycle, except for three [retroviruses (which go through a DNA intermediate) and the two RNA viruses that use the cellular splicing machinery (orthomyxo- and bornaviruses)]. All but four of the RNA viruses are enveloped—only the picornaviruses, astroviruses, caliciviruses, and reoviruses contain naked capsids. With the exception of the double-stranded RNA genomes of the reovirus family, all RNA virus families contain single-stranded genomes. These single-stranded RNA genomes can be of mRNA sense (i.e., positive sense), complimentary to mRNA (negative sense), or a combination of both (ambisense). Finally, most RNA viruses have nonsegmented genomes, with the exception of three families—orthomyxoviruses, bunyaviruses, and arenaviruses. An overview of the 15 families of RNA viruses based on their molecular properties and life cycles is presented below.

POSITIVE-SENSE RNA VIRUSES

There are six families of RNA viruses that contain a single-stranded genomic RNA molecule of message-sense. The genomic nucleic acid of positive-sense RNA viruses is fully infectious when delivered to cells in the absence of viral proteins. The gene expression strategies of these agents fall into three groups. The least complex is illustrated by the picornaviruses and flaviviruses. These agents have a single large open reading frame that encodes a single large polyprotein that is subsequently cleaved in an ordered fashion by proteases to generate individual functional polypeptides. Cap-independent translation was discovered in picornaviral internal ribosome entry site elements located in the 5′ noncoding region of the genome. Togaviruses, astroviruses, and caliciviruses generate two or more individual polyproteins that are encoded by the genome, as well as subgenomic mRNAs that are generated during infection. Coronaviruses use the most complex strategy. Instead of generating a polyprotein, numerous subgenomic mRNAs are made from their large genomes, each encoding an individual protein. Recombinant togaviruses that encode subgenomic mRNAs capable of producing large amounts of proteins of interest have been developed. Therefore in addition to naturally occurring positive-sense

RNA viruses, the possibility that designer viruses may be responsible for suspicious outbreaks cannot be overlooked.

Picornaviruses are a family of small stable viruses, most of which can survive the harsh environment of the gut and be recovered from stool samples. There are five key genera of picornaviruses. Poliovirus, a rather ineffective pathogen that can infect the central nervous system (CNS) and cause poliomyelitis, is eradicated in many parts of the world due to the use of highly effective killed and attenuated vaccines. Select members of the 30+ serotypes of echoviruses are among the most commonly isolated viruses from stools, and cause a variety of disease symptoms including meningitis, encephalitis, rashes, diarrhea, upper respiratory infections, and conjunctivitis. Coxsackieviruses (29 serotypes) are also rather commonly isolated from humans and cause hand, foot, and mouth disease (which is distinct from the animal virus that wreaked havoc on British livestock in 2000), herpangina, pleurodynia, and myocarditis in addition to a variety of nonspecific symptoms. Hepatitis A virus, the agent that causes acute infectious hepatitis, is also a member of the picornavirus family. The final branch of the human picornavirus family tree, rhinovirus, contains over 100 serotypes and causes the common cold. The biology of rhinoviruses is unique from other picornaviruses in that the virions are not stable to low pH, and the agent prefers growing at lower temperatures (33°C vs. 37°C). These properties are reflected in the restriction of rhinovirus infections to the upper respiratory tract.

Flaviviruses are enveloped viruses with icosahedral particles. There are six key members of this family, four of which are transmitted to humans via arthropod vectors. Yellow fever virus, the prototype of the group, was the first human virus isolated. It causes an *Aedes* mosquito-borne severe fever and myalgia, followed by jaundice and vomiting that can be prevented through the use of a live attenuated vaccine. The four serotypes of Dengue fever virus are also transmitted by *Aedes* species and cause a fever that can acquire hemorrhagic manifestations and be extremely severe. Dengue is probably the most significant mosquito-borne viral disease worldwide, with 50–100 million cases per year. St. Louis encephalitis virus is associated with *Culex*-borne epidemic encephalitis in the U.S. The last of the arthropod-borne flaviviruses, West Nile virus, is a key emerging pathogen in the U.S. since 1999. There are two hepatitis agents—hepatitis C virus (HCV) and hepatitis G virus (HGV)—that are classified as Flaviviridae that are not passed to human via an insect route. HCV is a major cause of chronic hepatitis in the U.S.

The four families of positive-sense RNA viruses that make subgenomic mRNAs contain species that cause a variety of human diseases. Key members of the Togaviridae include a variety of arthropod-borne encephalitis agents (Eastern equine encephalitis virus, Western equine encephalitis virus, and Venezuelan equine encephalitis virus) along with rubella virus (RV). RV is the agent that causes the fever and rash associated with German measles. More importantly, the virus can cause severe fetal abnormalities if it crosses the placenta. RV is currently under control in the U.S. due to the wide use of an attenuated viral vaccine given as part of the MMR vaccination. The Astroviridae and Caliciviridae are difficult to culture and are associated with gastroenteritis outbreaks. In addition, the hepatitis agent HEV (hepatitis E virus) is classified as a calicivirus. The final family of viruses in this group—the Coronaviruses, contains exceptionally large genomes (~27 kb) in a helical nucleocapsid and are associated with common colds. The agent responsible for the SARS outbreak in 2003 is a member of the Coronaviridae.

NEGATIVE-SENSE SINGLE-STRANDED RNA VIRUSES

All of the viruses in the four families of human pathogens classified in this group contain helical nucleocapsids within an enveloped particle. There are three general factors that govern gene expression in these viruses. First, the viral polymerase initiates transcription only at one end of the viral genome and reinitiates with approximately 50% efficiency following termination at the end of each gene. This property makes gene order an important means by which the relative levels of proteins are regulated. Second, the decision by the polymerase whether to transcribe mRNAs or replicate the entire genome is determined by the relative levels of the viral nucleocapsid protein. If levels are sufficient to package the nascent transcript, the polymerase ignores termination signals and proceeds copying the RNA template. Finally, cotranscriptional editing—the insertion of additional uncoded nucleotides by the viral polymerase—occurs at a select homopolymer run in all paramyxovirus genomes. This property is probably related to the tendency of the polymerase to stutter on poly(U) stretches to generate the poly(A) tail. In addition to the natural human viruses outlined below, it should also be noted that the technology to create recombinant negative-sense viruses has been developed.

The four families of negative-sense, single-stranded RNA viruses contain several classical and emerging human viruses of notoriety. The first family, the Paramyxoviridae, contains four major human viruses that all initiate infection in the upper respiratory tract. Measles and mumps viruses cause once-common childhood maladies (fever/rash and parotitis syndromes, respectively) that are preventable by live attenuated vaccines given as part of the MMR series. The four serotypes of human parainfluenza virus are responsible for numerous seasonal respiratory infections in children, most notably croup. The last member of the paramyxoviruses, respiratory syncytial virus, is the number one respiratory pathogen for children in terms of disease severity. No vaccine is currently available to prevent infection by the last two members of the paramyxoviruses. Prophylactic doses of anti-RSV IgG is currently being given to reduce to the risk of respiratory syncytial virus infection in highly susceptible groups of children. The major member of the next family, the Rhabdoviridae, is rabies virus. This bullet-shaped virion is usually acquired through the bite of a rabid animal (most notably bats in recent years). Following replication at the site of infection, the virus travels to the CNS, where it invariably results in death. Because of the protracted time it takes for the virus to cross the neuromuscular junction to infect the nervous system, this is the only viral infection where post-exposure immunization is helpful. The Ebola virus is the prototypic member of the third family of negative-sense RNA viruses, the Filoviridae. As its name implies, this family consists of viruses with elongated nucleocapsids. Ebola causes outbreaks of severe hemorrhagic fever that is often fatal. Although the disease has not been seen in the U.S., several individuals in Reston, Virginia were infected by an Ebola strain with reduced virulence that was carried in a primate laboratory animal shipment. The bornaviridae make up the final family of this group. These viruses are unique among the negative-sense RNA viruses in that they replicate in the nucleus in order to utilize the cellular splicing machinery. It has been suggested that bornaviruses play a role in several neural pathologies, but their exact contribution to human disease is still being explored.

SEGMENTED RNA VIRUSES AND THE RETROVIRIDAE

There are four families of human pathogenic RNA viruses that contain genomes consisting of two or more unique segments of nucleic acid. These genomes can take the form of negative-sense, ambisense, or double-stranded RNA. A key feature of these viruses is the rapid evolution that can occur due to reassortment of genomic segments in the progeny of infections by multiple viral species in the same cell. This is the reason that epidemiologists follow influenza virus very closely to insure that the current vaccine will provide coverage to any new variants that have recently emerged. One interesting molecular property of all of the single-stranded families in this group is that they “steal” the 5′ caps from cellular mRNAs to form the 5′ end of their own transcripts.

The segmented RNA virus families contain numerous important and interesting human pathogens. The Orthomyxoviridae consist of the enveloped negative-sense RNA viruses that cause influenza. Influenza A has eight segments. As noted above, due to reassortment, the nature of the two major surface antigens, hemagglutinin and neuraminidase, are closely monitored by the

U.S. Centers for Disease Control (CDC) in new influenza virus isolates. In addition to the killed influenza vaccine, infections can also be treated by antivirals that block uncoating (amantadine) and neuraminidase function. The Bunyaviridae family contains many arthropod-borne agents that cause fevers and encephalitis. These viruses have three genomic RNA segments that may be negative- or ambisense. In addition to these viruses, the rodent-borne hantavirus which emerged in the southwestern U.S. in 1993, causing numerous cases of severe pulmonary syndrome, is also a bunyavirus. The rodent-borne arenaviruses contain two ambisense genomic RNA segments and generally cause severe hemorrhagic fevers in humans (e.g., Lassa fever virus). The human members of the Reoviridae consist of 10–12 double-stranded RNA segments packaged in a double-walled protein shell. These viruses never fully uncoat their genome during infection, which helps avoid the action of double-strand activated enzymes of the interferon response. The major example of a human Reoviridae member is the common childhood diarrhea agent rotavirus. Over 80% of children in the U.S. will suffer diarrhea and vomiting due to infection by this agent.

The Retroviridae are included in this section not because they contain a segmented genome, but because they package two identical strands of positive-sense RNA. During infection, this genomic RNA is converted to DNA that then integrates into the cellular genome. Cellular DNA-dependent RNA polymerases are responsible for viral gene expression. The main examples of human retroviruses are the agent responsible for human acquired immune deficiency syndrome (AIDS: HIV) and human T-cell leukemia virus. Both of these agents are complex retroviruses that contain several regulatory genes (i.e., tat and rev) in addition to the gag, pol, and env genes of conventional retroviruses.

THE APPLICATION OF MOLECULAR FORENSICS AND EPIDEMIOLOGY TO VIRAL INFECTIONS

Antigenic subtyping has been applied for many years to classify viral infections and outbreaks. Elaborate networks of surveillance teams, for example, are monitoring the antigenic characteristics and epidemiology of influenza virus strains isolated from infected patients, to determine the best vaccine cocktail to use in the coming year. The advent of rapid sequencing techniques has opened the door to obtaining detailed fingerprints of viruses that could provide important clues as to their source. In general, a selected region of the viral genome is sequenced from numerous isolates and subjected to comparative phylogenetic analysis. Despite the increasing desire to rely on molecular analyses for all the answers, one should not overlook the fact that phenotypic features of the virus (tropisms, cytopathic effects, etc.) may provide important clues to facilitate identification.

As with other microbial species, the analysis of viral genomic segments can provide important clues as to the relatedness and origins of infections. Due to their small size and high rate of evolution, several considerations should be kept in mind when applying comparative molecular forensic analyses to viruses. First, for statistical reasons, it is usually advisable to look at as large a number of regions with variable sequence content as possible in the viruses that are isolated. Cost constraints, however, often make this impractical. Many viral genes, especially in RNA viruses, contain regions under rapid evolution and others that are under significant evolutionary constraint. In choosing a region of the viral genome to focus the analysis upon, it is important to consider that only rapidly evolving regions will provide enough useful variability when comparing agents from recently acquired infections. Second, if possible, it is extremely useful to include in the analysis a significant number of control viruses isolated from the surrounding population. This will allow for a full consideration of the background viruses in the local environment, and allow for a stronger statistical argument for relatedness to a predicted infection source. Finally, the overall strength of the argument for relatedness between two viruses requires detailed phylogenetic and statistical analyses that consider all alternative hypotheses. It will be important to get as close to "100%" certainty when one is performing the analysis for use in a legal rather than scientific context.

The tracking of HIV infections provides excellent examples of the successful application of molecular forensics to identify the source of a viral infection. In 1990, an HIV-positive dentist in Florida was suspected to be the source of HIV infection in six patients with no known risk factors. The sequencing of the env gene from viruses isolated from the doctor and patients strongly corroborated the epidemiological data that suggested the transmission route.[1] In more recent cases, polymerase chain reaction (PCR) amplification of the HIV genes followed by phylogenetic analyses have been used to suggest HIV transmission from a surgeon to a patient[2] and from a nurse to a patient.[3] Molecular forensics has been applied to other viral infections in addition to HIV. The nosocomial spread of specific strains of hepatitis C virus infections in hemodialysis units, for example, has been documented in several instances.[4,5]

The examples above involve molecular epidemiology of conventional viral pathogens. In order to rapidly identify viral agents involved in biocrimes, more work is needed to develop supportive resources. There is a clear need to establish an extensive sequence database of possible species. The identification of unique patterns and signatures beforehand will greatly facilitate the elaboration of the strain and perhaps the source of the suspect agent. In the case of recombinant biowarfare agents, the bioengineered features along with the strain background should give reasonable clues to the source.

In closing, the goal of this chapter is to provide a background in the fundamentals of human virology and provide an overview of the utility and issues surrounding the use of molecular forensics and epidemiology to the world of virology. The rapid evolution of genomic technologies should continue to expand the capacity and impact of this exciting field. I hope that this chapter can be used as a stepping-stone to appreciate and give perspective to these advances.

REFERENCES

1. Ou CY, Ciesielski CA, Myers G, Bandea CI, Luo CC, Korber BT, Mullins JI, Schochetman G, Berkelman RL, Economou AN, et al. Molecular epidemiology of HIV transmission in a dental practice. *Science* 1992;256:1165–1171.
2. Blanchard A, Ferris S, Chamaret S, Guetard D, Montagnier L. Molecular evidence for nosocomial transmission of human immunodeficiency virus from a surgeon to one of his patients. *J Virol* 1998;72:4537–4540.
3. Goujon CP, Schneider VM, Grofti J, Montigny J, Jeantils V, Astagneau P, Rozenbaum W, Lot F, Frocrain-Herchkovitch C, Delphin N, Le Gal F, Nicolas JC, Milinkovitch MC, Deny P. Phylogenetic analyses indicate an atypical nurse-to-patient transmission of human immunodeficiency virus type 1. *J Virol* 2000;74:2525–2532.
4. Seme K, Poljak M, Zuzec-Resek S, Debeljak M, Dovc P, Koren S. Molecular evidence for nosocomial spread of two different hepatitis C virus strains in one hemodialysis unit. *Nephron* 1997;77:273–278.
5. Halfon P, Roubicek C, Gerolami V, Quentin Y, Khiri H, Pepe G, Berland Y. Use of phylogenetic analysis of hepatitis C virus (HCV) hypervariable region 1 sequences to trace an outbreak of HCV in an autodialysis unit. *J Clin Micro* 2002;40:1541–1545.

CHAPTER 4

Keeping Track of Viruses

JACK HIETPAS, LAURA K. MCMULLAN, HOLLY L. HANSON,
AND CHARLES M. RICE

Laboratory of Virology and Infectious Disease, Center for the Study of Hepatitis C, The Rockefeller University, New York, New York

DAVID P. MINDELL

Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, Ann Arbor, Michigan

INTRODUCTION

Viruses are ubiquitous in nature. They prey on all forms of life from archaea and eubacteria to fungi, plants, and animals. It has been estimated that viruses that target sea-dwelling organisms, such as algae, turn over half of the world's biomass each day. As such, viruses are an important part of the ecological balance on earth. Of the vast collection of ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) viruses, research has focused on the tiny fraction that cause disease in humans, domestic animals, and crops. Modern virology, coupled with public health initiatives, has seen many triumphs including the development of effective vaccines against yellow fever, poliomyelitis, smallpox, mumps, measles, rubella, hepatitis A and B, and papillomavirus-associated cervical cancer. The global eradication of smallpox and poliovirus as well as the prediction and successful prophylactic vaccination against influenza represent tremendous gains for public health. Antiviral therapy against human immunodeficiency virus (HIV) and some herpesviruses, as well as rapid detection and containment of emerging viruses such as the severe acute respiratory syndrome coronavirus (SARS-CoV) illustrate the successful application of viral research programs as well as the need for ongoing efforts focused on rapid identification and elucidation of viral distribution, kinetics, and pathogenesis. With the growth of knowledge about replication, transmission, and disease, viruses have also taken the stage as possible biowarfare, bioterrorist, and criminal agents.

Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

In our unpredictable world, viral diagnostics and forensics are becoming increasingly important and sophisticated. The purpose of this chapter is to introduce basic concepts in virology and to provide several examples of how current technology has been used to identify and track viral pathogens. The first part of the chapter discusses virus structure, replication strategies, classification, and evolution. The second part highlights the varied modes of transmission, infection, and disease manifestations as they relate to the different diagnostic methods for virus identification. We follow this with examples of how emerging viruses were identified and tracked (Sin Nombre, Nipah, West Nile, SARS coronavirus, and monkeypox). With regard to viral forensics, the Schmidt case is a fascinating example illustrating how the study of RNA virus evolution and relatedness was used in a criminal investigation. Finally, we provide a few snapshot views into new technologies that allow the creation of manmade or “engineered” viruses, some exciting new frontiers in viral diagnostics, and a few thoughts on the future of virology.

WHAT IS A VIRUS? BASIC VIROLOGY

VIRUS ANATOMY

Viruses are extremely simple “life” forms without metabolic capacity, organelles, translational machinery, or autonomous replicative potential; the essential elements of a virus are consequently minimal. Virus particles are either enveloped or nonenveloped. Enveloped viruses contain one or more envelope glycoproteins embedded in a lipid bilayer that is acquired by budding through host cell membranes. Nonenveloped viruses are made up of a tight protein shell. In both kinds of viral particles, the envelope or protein shell serves to protect the viral genome from the hazardous extracellular environment, since without such protection, the fragile nucleic acid genome would be vulnerable to physical, chemical, and/or enzymatic destruction. Surface proteins also harbor functions required for virus entry into host cells. Inside the virus particle, the viral genome is associated with one or more proteins to form a complex usually referred to as the nucleocapsid. In addition to nucleic acid binding proteins, some viruses also carry enzymes required to initiate replication (like polymerases). Although not all virus particles are highly structured, two kinds of symmetry are common. Helical symmetry is reflected by the arrangement of the protein subunits in a “spring-like” or stacked lockwasher fashion, whereas icosahedral symmetry comprises 20 triangular faces that form the surface of a sphere. In the simplest icosahedral capsid, one protein subunit is at each triangular point of the 20 faces, requiring 60 total subunits.

While host organisms use only DNA for their genetic material, a viral genome may be composed of either DNA or RNA. The size of viral genomes varies greatly. The genome of variola major, the DNA virus that causes smallpox, is 190 kb. The RNA genome of the SARS-causing coronavirus is 29.7 kb; those of Ebola and Marburg viruses are each 19 kb, while the HIV and poliovirus are 9.2 kb and 7.4 kb, respectively.

VIRUS LIFECYCLE

VIRUS ATTACHMENT AND ENTRY

Among the enormous variety of host cell types, how does a virus choose which cell to commandeer for its own replication and propagation? The surface of an enveloped virus particle contains virus-specific proteins that mediate viral recognition of host cell targets. The membrane surface of a host cell is riddled with numerous macromolecules including glycoproteins, glycolipids, and carbohydrate residues that have specific host functions serving as transport channels, mediating signaling pathways, and assisting in antigen presentation. While these macromolecules may be essential for host cell survival and function, some of these macromolecules also serve as receptors for virus particles. The presence of the host cell receptor is essential for viral attachment, but the mere presence of the receptor does not confer upon the cell the ability to support virus replication. Some viruses (such as HIV) require a coreceptor for viral entry without which the virus may be able to bind to the host cell membrane but remain unable to enter the cell. If the virus is not enveloped or is a naked particle, the viral capsid undergoes a conformational change upon binding that releases energy used to breach the host membrane, allowing viral entry. Delivery of viral nucleocapsid, and hence the genome RNA or DNA, occurs via fusion of the virus particle with cellular membrane either at the cell surface or after endocytosis. Virus genome replication can occur either in the cytoplasm or the nucleus, and all viruses usurp the host cell translational machinery to produce virus-specific proteins.

REPLICATION STRATEGIES

In 1971, David Baltimore divided viruses into seven groups based on genetic material, polarity, and mRNA synthesis.[1] Insight into the replication strategy of a virus can be gained by understanding how other well-studied viruses in the same group replicate. The following is a brief description of the seven groups; included where possible, are examples of viruses belonging to each

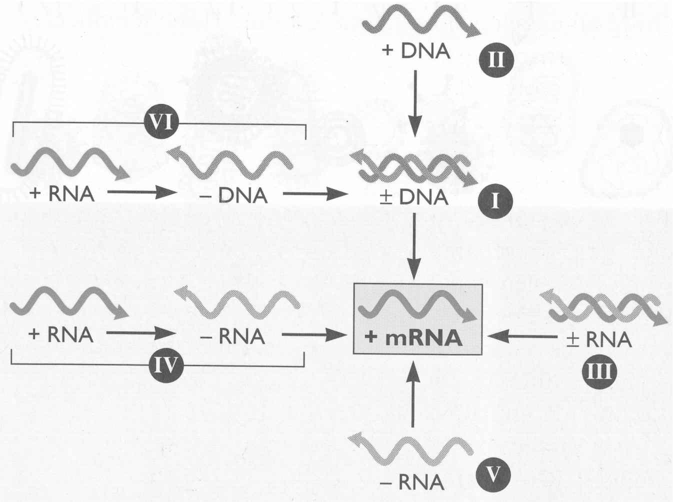


FIGURE 4.1 Schematic illustrating the replication strategies and means of gene expression for the various Baltimore classes. Reprinted with permission from Ref. (56). (See color insert.)

given group. Note that in the following sections nucleic acid genomes will be discussed in terms of strand polarity, where positive, (+), strand nucleic acid has a 5′→3′ polarity identical to mRNA and negative, (–), strand nucleic acid is complementary to mRNA. In cases of (–) strand polarity, the virus must synthesize a complementary (+) strand for gene expression to occur. A simplified schematic is presented in Figure 4.1.

Double-stranded DNA viruses (Class I): Viruses of this class have double-stranded DNA genomes and are subdivided into two groups. The first contains viruses that require replication to take place in the nucleus; the second class contains viruses that replicate in the cytoplasm. Variola major, the cause of smallpox, is an example of an enveloped double-stranded DNA virus that replicates in the cytoplasm; its linear DNA genome is directly transcribed to mRNA, which is then translated by host machinery to produce viral proteins. In contrast, genome replication of herpesviruses occurs in the nucleus.

Single-stranded DNA viruses (Class II): These viruses have single-stranded DNA genomes. Host proteins transcribe mRNAs from the viral genome, which are subsequently translated to viral proteins.

Double-stranded RNA viruses (Class III): Viruses in this class contain double-stranded, segmented RNA genomes. mRNA is synthesized by a virally encoded RNA-dependent RNA polymerase (RdRp) that is, like the RNA genome, contained within the capsid. Most eukaryotic cells do not encode RdRps, so in order for the virus to replicate itself, it must provide this enzyme for RNA replication. The common etiologic agent causing severe infectious diarrhea in children, rotavirus, has a double-stranded, segmented RNA genome.

Positive-strand RNA viruses (Class IV): These viruses have single-stranded (+)-strand RNA that is directly translated by the host cell to produce viral proteins. The (+)-strand genomic RNA from these viruses can be infectious, in contrast to (−)-strand RNA, which cannot. To generate multiple copies of the genome, these viruses synthesize (−)-strand complementary RNA species that are subsequently transcribed by viral RdRps, to produce more (+) strands. Examples of viruses in this class include poliovirus, West Nile virus, the SARS coronavirus, and hepatitis A virus.

Negative-strand RNA viruses (Class V): In contrast to class IV, viruses of this group contain negative polarity single-stranded RNA molecules as their genome. These viruses are all enveloped and can have genomes that are either segmented or continuous. Some members have ambisense genomes with portions of the genome acting as (+) strands and other portions of the genome having (−) polarity. All members of this class, such as influenza virus, hantavirus, and Ebola virus package both genome RNA and an RdRp into their virion.

Retroviruses (Class VI): This unique class of viruses uses a totally novel scheme for replication and expression. These viruses have two identical copies of single-stranded (+)-polarity RNA molecules as their genome. These RNA molecules are reverse transcribed by the enzyme reverse transcriptase (RT), generating complementary DNA molecules from their RNA templates. Members of this class are called retroviruses, reflecting the fact that their replicative cycle is retrograde (RNA→DNA→mRNA→protein) relative to the central dogma of modern biology, in which the DNA is transcribed to mRNA, which is then translated to protein (DNA→mRNA→protein). A very important virus in this class is the human immunodeficiency virus (HIV), which causes acquired immune deficiency syndrome (AIDS).

Hepadnaviruses (Class VII): Members of this group have partially double-stranded DNA molecules as their genome. These viruses replicate via an RNA intermediate, similar to the retroviruses. RNA is packaged into immature particles where reverse transcriptase uses the RNA template to generate the DNA genome. An example is hepatitis B virus, an important human pathogen that can cause chronic infection and consequent liver damage.

ASSEMBLY AND RELEASE

Once the genome of a virus has been amplified, it must be packaged into infectious particles. This is a complicated process that is well understood for only a few viruses. The exact mechanisms that allow selective packaging of viral nucleic acid and particular virus-specific proteins are not completely elucidated. The site of assembly can be in the nucleus, in the cytoplasm in association with membranous organelles (endoplasmic reticulum or Golgi bodies), or at the cell surface. The process of virion release occurs through budding or in the case of naked nonenveloped viruses, through host cell lysis.

VIRUS CLASSIFICATION

Viral nucleic acid, replication mechanisms, and virion release are only some of the characteristics that define a virus. Virologists have devised classification systems to describe and identify viruses based on observed similarities and differences. Initially, the decision of how to classify a virus was rooted in understanding the host and tissue infected by the virus, the resultant disease process, and the method of transmission. As technology advanced, the classification criteria shifted to include physical properties such as virion morphology, stability, filterability, and antigenicity. With the development of negative-staining electron microscopy that allowed scientists to visualize the virus particle, the classification of viruses changed to comprise descriptions of virion size, morphology, and surface characteristics. Finally, technological advances including polymerase chain reaction (PCR), DNA sequencing, monoclonal antibody production, and DNA microchip analysis have allowed researchers to differentiate between single-nucleotide changes present in viral genomes. The complete genomes of numerous viruses have been sequenced, allowing for the identification of species, strains, and mutants that could never before have been possible. When very little information is known about an emerging infection, the classification system and its wealth of descriptive information is useful; predicting possible routes of transmission and pathogenic consequences based on information known about similar well-studied and previously classified viruses may significantly impact public health practices while the new etiologic agent is under study.

The current standardized hierarchical International Committee on the Taxonomy of Viruses (ICTV) system used to classify viruses utilizes sequentially more exclusive levels of order, family, subfamily, genus, and species.[2] This classification system includes viruses that infect vertebrates, invertebrates, bacteria, fungi, plants, algae, and others. A brief description of the classification system follows.

VIRUS ORDER

Virus order is a collection of phylogenetically related virus families that have similar properties. At present there are only two recognized orders, *Monone-*

gavirales and *Nidovirales*. The *Mononegavirales* order is comprised of viruses that have a single-stranded negative-sense RNA molecule as their genome. There are three families in the order *Mononegavirales*: *Filoviridae*, *Rhabdoviridae*, and *Paramyxoviridae*. The *Nidovirales* order comprises the families *Coronaviridae* and *Arteriviridae*. The virus that causes SARS is a member of the *Coronaviridae*. Virus order names are discerned from the other classification level names by ending with the suffix *-virales*.

VIRUS FAMILY/SUBFAMILY

A virus family is a phylogenetically related collection of genera that share similar properties. Families are given names that end with the suffix *-viridae*. Virus particle morphology, genome characteristics, and/or strategies of replication are used to group viruses into families. Due to the diversity and complexity of the families *Paramyxoviridae*, *Poxviridae*, *Parvoviridae*, and *Herpesviridae*, each is further classified into subfamilies. Subfamilies are given names that end in *-virinae*.

VIRUS GENUS

Names given to genera end with the suffix *-virus*. There is continuing pressure to use increasingly detailed structural, physicochemical, or serological differences to create new genera in many families.[3]

Virus Species

Species is a highly specific classification given to a virus that has met many selective criteria. In 1991, the ICTV defined a species as "a polythetic class of viruses that constitutes a replicating lineage and occupies a particular niche."[3] This definition allows for some plasticity in the properties required for species classification, allowing emphasis in some cases to be on genome properties, in others on structural, physicochemical, or serological properties. The term "quasi-species" refers to a population of closely related viral sequences. This is often used to describe RNA virus isolates where essentially every genome sequence is different due to the high error rate of RNA-dependent RNA replication.

The highly structured and formal classification system set up by the ICTV gives the virologist, clinician, technician, epidemiologist, and first-responder access to valuable information. Once a virus is identified and its classification is defined, information about the virus or similar better-studied viruses such as the kinetic characteristics of virus spread, tissue tropism, disease pathogenesis, associated clinical symptoms, treatment, and level of required biocontainment can be quickly realized. This system also allows for classification and characterization of a new unknown virus. For example, SARS was a previously unknown virus that was quickly analyzed and placed into the *Coronaviridae* family.

GENERATION OF VIRAL DIVERSITY

Although the classification system facilitates virus identification, the constant emergence of new, diverse viral populations continues to challenge scientists. Three major mechanisms give rise to viral diversity: mutation, reassortment, and recombination. Depending upon the replication strategy of a particular virus, the relative contribution of these pathways differs. Nucleotide misincorporation leading to mutation occurs for all viruses but is dramatically different for viruses with RNA versus DNA intermediates. Errors are made both by DNA and RNA polymerases, but while DNA polymerases have proofreading capacity, RNA polymerases do not. Consequently, errors generated by RNA polymerases are incorporated into the newly generated viral genome at a very high rate. The error rates of DNA polymerases can be as low as one base substitution per every 10^8–10^{11} nucleotides, whereas the RNA polymerases have error rates as high as one base substitution per every 10^3–10^6 nucleotides.[4] In the case of RNA viruses, virus population studies of patient and tissue culture isolates have shown the coexistence of numerous viral genomes with slightly different sequences, or quasi-species.

Homologous or nonhomologous recombination is another mechanism by which new viruses are generated and has played a significant role in virus evolution. Large DNA viruses can not only recombine with each other but also contain genes usurped from the host. RNA tumor viruses have incorporated cellular oncogenes that lead to transformation in some cell types. For RNA viruses, recombination probably occurs via incomplete synthesis and template RNA switching during negative-strand synthesis. This process occurs with high frequency in poliovirus infection, with an estimated 10%–20% of poliovirus genomic RNA recombining in a single growth cycle.[5] In polio eradication efforts, a live virus preparation consisting of three separate attenuated poliovirus strains was used. Despite the safety of each attenuated strain, some vaccinated patients did develop poliomyelitis. Through genetic testing of patient isolates, it became clear that the recovered virus was not the same as the administered vaccine, and that neurovirulence had been restored by recombination among the three attenuated strains.[6–7]

Reassortment can also occur when two related viruses with segmented genomes infect a single cell. For example, reassortment is responsible for generating new influenza A strains that are antigenically different enough from previous strains to infect and cause disease. Besides the annual flu season, severe pandemics have occurred such as the Spanish flu of 1918, the Asian flu of 1957, the Hong Kong flu of 1968, and the Russian flu of 1977.[8] Influenza virus infects several animal hosts, including humans, pigs, horses, and its natural reservoir, waterfowl. The two major antigenic proteins of this virus, hemagglutinin (HA) and neuraminidase (NP), can be reassorted between different influenza viruses, generating new strains different from the two parental strains (called antigenic shift). Each influenza strain is given a code which identifies its HA and NP proteins. The Spanish flu was caused by the H1N1 strain, the Asian flu was strain H2N2, the Hong Kong flu was strain H3N2, and the Russian flu was strain H1N1.[9]

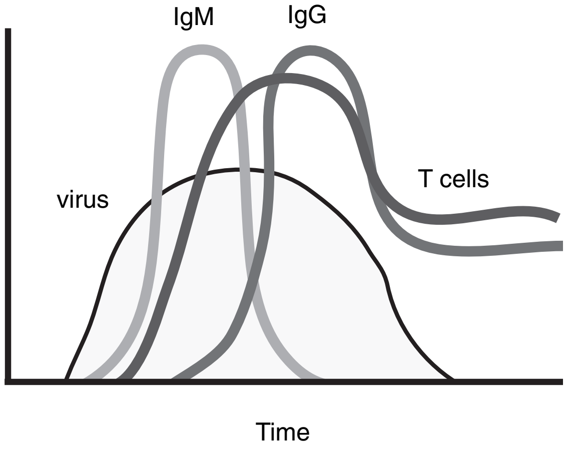
INTRODUCTION TO VIRAL KINETICS AND OUTCOME

A virus enters the host through any opportunistic opening of the body. Minute breaks in the skin barrier or mucosal linings of the eye, nose, gastrointestinal, and urogenital tracts create portals for the virus to enter host cells and underlying tissues. Replicating virus may establish a localized infection, or the virus, traveling as free virions or associated with immune cells, may spread to other areas of the body. Replication can continue at the initial site of infection or move to secondary organs and tissues. During an active infection, virus particles may be shed through nasal and respiratory secretions, urine, and stool.

INCUBATION PERIOD

An incubation period is the time between exposure to virus and the manifestation of symptoms of infection. Initial symptoms may be slight, such as a rash or fever, and may not initially prompt patients to seek medical care. Diagnostic samples are taken when people seek medical treatment; therefore, the presence of virus and virus-specific antibodies depends on the incubation period. This observation has important implications for the identification of new viral infections or epidemics. Virus may be detected and isolated for infections with short incubations; however, antibodies may not yet have developed, see Figure 4.2. For example, infections with influenza and adenovirus have very short incubation periods, 1–2 days, before myalgias and fever manifest. Many viruses have invaded the body for over a week before symptoms become apparent. Dengue, measles, and rotavirus are relatively asymptomatic during the first 5–8 days, but may rapidly progress to severe illness soon after. Longer incubation

A. Acute infection



B. Chronic Infection

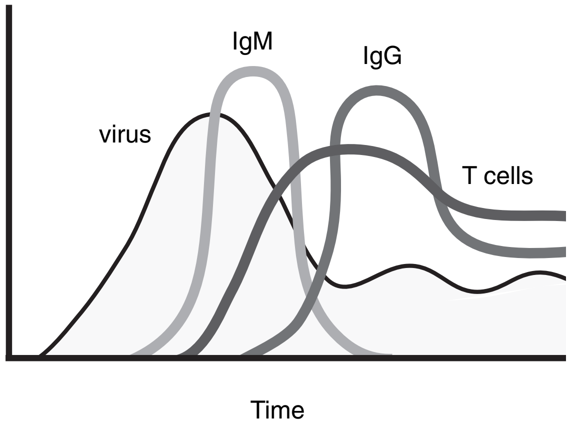


FIGURE 4.2 (A) Adaptive immune response in an acute resolving infection. Low-affinity antibodies of the IgM isotype develop initially, and partially control circulating virus. T cells undergo clonal expansion upon virus antigen stimulation. In response to effector T cells and often higher-affinity IgG antibodies, virus levels decline and the immune system differentiates into memory T cells. Such memory T cells and neutralizing antibody may persist for the life of the patient. (B) Adaptive immune response in a chronic virus infection. Antibodies and T cells develop but are unable to eliminate virus. Virus levels, although typically lower than in the acute stage, are continuously present despite a measurable immune response. Mechanisms leading to chronic infection include generation of viral mutants that escape the adaptive immune response, induction of immune exhaustion or tolerance, and a quiescent state called latency for some DNA and RNA viruses.

periods may be caused by chronic infections. For example, rabies virus may remain silent for 30–100 days after an animal bite before causing sudden disease demanding medical treatment. Other stealth viruses, such as HIV, hepatitis B and C, and papilloma viruses, may bring about a mild ailment initially that may gradually progress to a fatal disease. Patients with chronic infections may seroconvert and have detectable virus even in the presence of antibodies. Others may be persistently infected by a virus, yet never develop symptoms. These carriers may have low levels of virus, which could potentially be transmitted to other individuals despite being below the level of assay detection. An understanding of how the host responds to viral infection aids investigators and health care workers in identifying and treating infection.

THE IMMUNE RESPONSE

The immune system typically recognizes the invading pathogen at the site of entry, where it launches a generalized innate immune response. Signal molecules recruit inflammatory cells to the area which engulf cell debris and released virions. Virus proteins are presented to the adaptive immune system in local lymph nodes, allowing the host to create a directed defense against the virus. Effector T cells specific for virus antigen stimulate B cells to secrete virus-specific antibodies. There are two predominate types of antibodies produced in response to virus infection, IgM and IgG. The first antibody produced is the IgM isotype and is present in the blood within days and remains in circulation for several months. As the adaptive response develops, B cells switch antibody isotypes to secrete the IgG form. Investigators use the detection of antibodies as indicators of virus exposure, and can utilize knowledge of the presence of specific isotypes to discern acute infection from past or chronic infection. Detection of IgM or rising titers of IgG points to a recent virus exposure. Virus-specific antibodies are also used as tools in numerous serological assays. The specificity of the antibody is useful to distinguish between different virus types and to detect virus protein in clinical samples.

COURSE OF INFECTION

Most virus infections cause an acute infection with viral particles rapidly cleared from the body. Examples of acute infections are influenza, rhinoviruses, and rotaviruses where infections are brief and virus is eliminated completely from the host. Other viruses are capable of establishing a long-term, persistent infection and may be present for years or throughout the life of the host. During a persistent infection, some viruses are continuously present in the body with a constant level of viremia, or virus detectable in the blood. For example, lymphocytic choriomeningitis virus is noncytopathic even while actively replicating. Other viruses such as Epstein-Barr virus regulate their replication programs to avoid detection by the immune system, while still others establish infection in certain areas of the body that are immunoprivileged sites, including the central nervous system (CNS) and the eye, where once present, viruses may propagate undetected. Viruses may evade the immune response by infecting the immune cells themselves, such as in HIV infection of $CD4^+$ T cells and dengue virus infection of dendritic cells. Some viruses enter a latent stage where replication temporarily ceases. Reactivation of virus replication can occur weeks to years later, multiple times in the life of the host. Finally, viruses may reduce the rate of virus production so that little or no virus is detectable. As the immune system struggles to control, but not eliminate the infection, virus may be measured in a cyclical pattern with titers alternately rising above and falling below the threshold of detection.

HOW DO YOU IDENTIFY A VIRUS?

An understanding of the established classification system, mechanisms responsible for generating viral diversity, and pathogenesis facilitates the identification of an unknown virus. The sudden emergence of an infectious disease demands rapid methods to identify the source of infection, diagnose patients, and explain routes of transmission. Detection of a novel virus relies on adequate sample acquisition and knowledge gained from previously identified viruses, sequences, and reagents such as antibodies and virus antigen.

SITES FOR VIRUS ISOLATION

The types of samples and the manner of collection depend upon the capabilities of the diagnostic laboratory. The chance of obtaining an active virus sample depends on many factors. The virus concentration varies in the body over the course of infection and may actually not be present when symptoms bring the patient to medical attention. The virus may have been cleared from the area of disease, or the symptoms may reflect damage induced by the immune system rather than direct virus pathology. It is therefore critical to obtain samples early in the infection and from multiple locations in the body. The most common site to look for virus is in the blood. Other clinical samples are generally collected near the site of virus-induced disease or based on known routes of virus transmission. Mosquitoes can deposit West Nile virus and dengue into the skin; poxvirus and papillomavirus replication can create lesions and warts; the perfuse spread of Ebola allows detection even in the skin. For respiratory infections, samples may include nasal and throat swabs as well as nasopharyngeal aspirates and bronchial lavage fluids. Urine and stool samples are collected for enteric disease, and cerebral spinal fluid is obtained when neurological symptoms are present. Viruses may be very labile, and care must be taken to avoid exposure to harsh treatments such as extreme pH, direct sunlight, and freezing temperatures in order to preserve the potential for identification and study. Generally, samples are transported on ice and evaluated in the lab as quickly as possible. If delays are expected, samples can be frozen, but the recovery of infectious virus may be reduced or eliminated.

VIRUS ISOLATION

Historically, viruses were cultivated in the laboratory utilizing human samples to directly inoculate susceptible animals or embryonic chicken eggs. Tissue culture allows virus to be isolated, quantified, and amplified to produce a large stock for further analysis. Primary cell lines derived from animal tissues have a limited lifespan in tissue culture, yet are necessary to initially isolate some viruses. Continuously replicating cell lines are the most convenient, with an infinite capacity to be passaged. Human specimens are clarified and treated with antibiotics before addition to cell culture. The inoculated cells are monitored for any detectable morphologic changes and periodically harvested for other serological assays. There are specific culture conditions for particular viruses. For instance, viruses from the conjunctiva, nasal and respiratory tracts are incubated at cooler temperatures resembling the exterior of the body, and viruses entering cells via fusion proteins requiring cleavage need to have trypsin added to the culture media. Some viruses cannot be cultivated in cell lines but require that clinical samples be injected into suckling mice by specially trained personnel and examined for virus using serological methods. Other culture conditions are being established that involve the use of whole-organ cultivation. Some viruses are even more difficult; hepatitis C virus can only be grown in chimpanzees.

ELECTRON MICROSCOPY FOR VIRUS IDENTIFICATION

In the early days of virology, researchers and clinicians could only identify and classify a virus based on its infectivity and associated pathology. Viruses could not be more closely described because of their extremely small size; virion particles can range from tens to several hundred nanometers (nm). The light microscope, which is a valuable instrument for the identification and study of fungi, bacteria, mold, and spores, does not have the resolving capacity to visualize virus particles. Virologists were limited in their investigations until the 1930's, when the electron microscope (EM) was invented. This instrument uses, instead of visible light, a beam of electrons to form an image of the specimen. The instrument not only produces extremely large magnifications (up to 1,000,000×) but also resolves fine structure at such magnifications.

Virus morphology is as diverse as that seen in other microscopic organisms. Because of this, viruses can be identified based on their appearance under the microscope. The researcher is able to examine both the external and internal structure of the virus. Particle morphology has been observed to appear ribbon-like (rabies), rod-shaped (measles), spherical (poliovirus), and filamentous (Ebola) (Figure 4.3). Many viruses show multiple morphologies under the microscope and are referred to as pleomorphic. An example of a very pleomorphic virus is influenza A, which can appear kidney bean-shaped or filamentous. Some viruses have very characteristic spikes, club or pin-like projections present on the viral envelope. A very interesting example of spikes is seen in the emerging SARS virus.

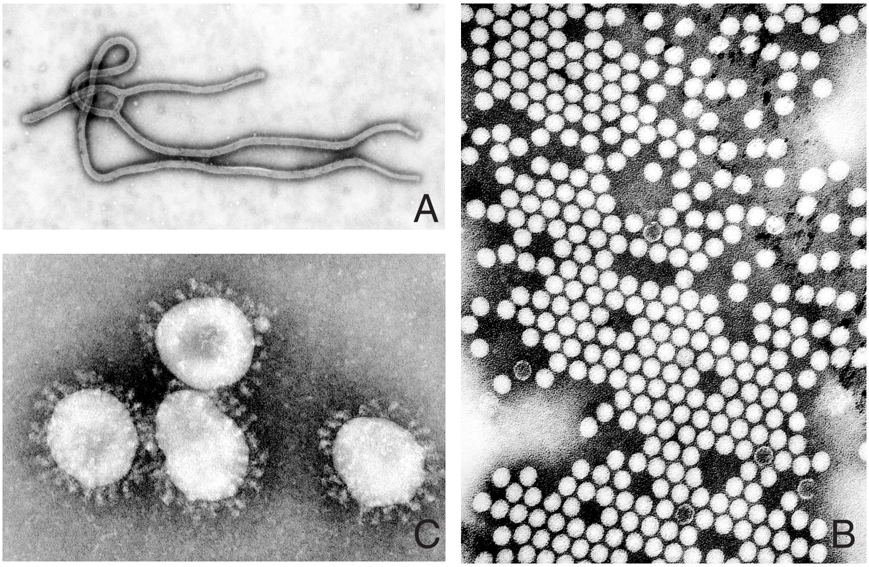


FIGURE 4.3 (A) Ebola virus particles showing filamentous morphology. Courtesy of the CDC/C. Goldsmith. (B) Polio virus particles showing spherical morphology. Courtesy of the CDC/Fred Murphy, Sylvia Whitfield. (C) SARS virus showing indicative "corona-like" morphology. Courtesy of the CDC/Fred Murphy.

There are two main types of visualization for virus identification using EM. The first is negative staining/contrast, which was a revolutionary visualization technique designed in the late 1950s. Due to the low level of electron scatter by viral particles, direct contrast and visualization are difficult. However, using the negative contrast technique, the virus particle is visualized on a black (electron-opaque) background. The degree to which the heavy metal salt stain (electron-rich) penetrates into the virus particle determines the contrast/resolution of virus structure. This procedure provides very useful information on the external structure of the virus, but is somewhat limited in resolving internal structure.

Thin sectioning is a very important sample preparation technique that allows internal structure to be visualized. Tissue samples are embedded with an epoxy-resin, and thin slices are prepared using an ultramicrotome. These slices are then examined using various electron-dense stains and antibodies that can be used to tag specific cellular organelles and virus proteins.

OBSERVANCE OF CYTOPATHIC EFFECT

A virus may kill an infected cell creating characteristic cytopathic damage or may replicate in cells without any visible effect. The type of changes induced by a virus can be significant for the type of virus and can be observed by the light microscope. The cytopathic effect (CPE) can be focal, diffuse through the cell monolayer, or at the edge of the culture. Cells may appear to be rounded or enlarged, growing in grapelike clusters, indicating adenovirus or herpes simplex virus. Influenza and mumps cause cells to fuse together, creating syncytia that detach from the surface. Cells may fuse and form multinucleated giant cells with granular cytoplasm typical of measles infection. Vaccinia and poxviruses create foci of fused cells, whereas picornaviruses induce proliferation of membranes in the cytoplasm and shrinkage of the nuclei (pyknosis). However, many viruses such as members of *Bunyaviridae*, *Arenaviridae*, and *Retroviridae* fail to produce obvious CPE and can replicate in culture without any noticeable change.

SEROLOGY

The specific affinity of antibody-antigen recognition is widely used in virus diagnostics. Polyclonal and monoclonal antibodies can be raised against recombinant virus proteins and virus produced from infected cells. If a virus can be isolated in cell culture, cell lysates and slurries can be used as virus antigens. Infected cell cultures can also be used to test for antiviral reactivity present in patient serum. Recombinant virus proteins produced in *E. coli* and baculovirus expression systems are also used as antigens to detect antibodies present in human serum, cerebrospinal fluid (CSF), and tissues. Finally, convalescent serum proven to react with infected cells in culture can be used as a source of antibodies for additional assays.

NEUTRALIZATION ASSAY

Antibodies produced during an infection often have the ability to bind the virus and reduce infectivity. These protective, neutralizing antibodies recognize epitopes on the surface of the virus and prevent virus from infecting a cell. Neutralizing antibodies are often used to classify virus into serogroups. Closely related virus families have similar virus coats, and neutralizing antibodies that can bind viruses within a group are said to cross-react. In a neutralization assay, dilutions of neutralizing antibodies are mixed with virus and assayed for remaining infectivity. An unknown virus may cross-react with antibodies from a known serogroup, revealing an antigenic relationship and an initial clue to the genetic identification. More specific and focused assays can then be performed to identify the virus. An extensive bank of serotype-specific antibodies in the diagnostic laboratory increases the chance of discovering a novel virus.

HEMAGGLUTINATION ASSAY

The hemagglutination assay uses the ability of virus proteins to bind and aggregate erythrocytes. Virus is mixed with an erythrocyte suspension in serial dilutions and added to a microtiter plate with V-bottom wells. Unabsorbed red blood cells fall to the bottom point of the well forming a dot, whereas aggregated blood cells uniformly coat the well. The hemagglutination inhibition assay is used to classify virus families that share that ability to bind to erythrocytes. Antibodies specific for a virus family may prevent aggregation of erythrocytes and are assayed by hemagglutination inhibition.

COMPLEMENT FIXATION

Complement fixation uses known virus-specific antigen rather than known anti-virus antibodies. Interactions between virus antigen and antibody cause fixation of complement and result in membrane lysis. Briefly, patient sera are incubated with antigen and a standardized amount of complement. Red blood cells are coated with anti-red blood cell antibodies that are recognized by complement and added to the sample. If antibodies specific to the virus antigen are present and fix complement the red blood cells will be protected from lysis. Using group-specific virus antigens allows a patient's virus to be identified.

IMMUNOSTAINING

Antibodies can be used to detect virus antigen in patient tissues and in infected cell cultures. An antibody is linked to a fluorescent dye or enzyme and allowed to bind virus antigen. The complex is visualized under a microscope, allowing location of the virus proteins to be seen inside the cell. Cross-reactive sera can be used to serologically identify an unknown virus as a member of a known family. Polyclonal sera can identify a wide range of viruses, whereas monoclonal antisera can give a specific diagnosis. Additionally, the direct visualization of virus antigen in a cell can provide interesting clues of virus properties beyond the diagnosis.

ENZYME-LINKED IMMUNOSORBENT ASSAY

The enzyme-linked immunosorbent assay (ELISA) provides a fast method to detect the presence of virus antigen or antibody from a large number of samples. Because of its reproducibility and flexibility, it is the most common serological assay used in the viral diagnostics lab. The general scheme is to capture virus antigens or virus-specific antibodies on a solid surface and to expose the bound complex with a substrate. ELISAs specific for IgM and IgG antibodies can distinguish between recent infection and previous exposure or vaccination. Detection of virus antigen indicates an acute stage of virus infection. The microtiter plate format can test multiple patient samples with different dilutions of antibodies to determine a serological titer, and sequential sampling of patient's serum over the course of an acute infection can allow detection of seroconversion. An increase in titers of IgG antibodies twofold to fourfold over 2 weeks is diagnostic of an active primary infection. ELISAs are frequently used in serosurveys where populations are screened to determine the efficacy of immunization programs or previous exposure to a newly identified virus.

NUCLEIC ACID-BASED METHODS

The detection of virus nucleic acid has revolutionized diagnostics. The use of molecular techniques has identified viruses that cannot infect cells in culture or inoculated animals. Most specimens such as blood, tissues, urine, stool, CSF, and respiratory secretions can be treated and the nucleic acid extracted, inactivating the virus and reducing concern about transport or contact in the laboratory. Detection of virus in the blood is diagnostic of an active infection, and the sensitivity of PCR has enhanced our capability to detect very low levels of virus in persistent infections; the universal genetic code allows investigators to detect virus sequence from host insects, animals, bacteria, and fungi where serological reagents may not be available. The specificity of PCR is dependent on the selection of the primer sets. Primers are designed to anneal to conserved regions in a virus genome. Genes that encode proteins essential for the virus life cycle such as polymerases, helicases, and integrases as well as RNA elements in untranslated regions are maintained even in viruses with high mutation rates. There are good PCR amplification targets. To identify a new virus, universal primer sets are designed based on alignments of virus family sequences and selected to minimize base pair mismatches. Primers containing degenerate bases are used to widen the net and find a genetically distant virus. For DNA viruses, PCR can be applied directly. RNA viruses require an initial reverse transcription RT-PCR reaction before amplification. Complex variations on the basic technique of PCR including utilization of multiplex primers, TaqMan probes, and molecular beacons can enhance the ability to identify viral sequences.

PCR products are either sequenced directly or cloned into intermediate vectors for sequencing. Multiple PCR isolates are sequenced in order to verify the authentic virus sequence, as mutations arise during PCR amplification. Sequencing virus directly from clinical samples without amplification in cell culture minimizes cell culture adaptation artifacts. The unknown virus sequences are entered into sequence databases to search for any similarity with known virus genomes. The database should include sequences from clinical isolates, field samples from the natural hosts, and nonpathogenic viruses to assist identification of a novel emerging virus. As with any assay, molecular diagnosis of virus nucleic acid is confirmed with other diagnostic methods such as serology and virus isolation.

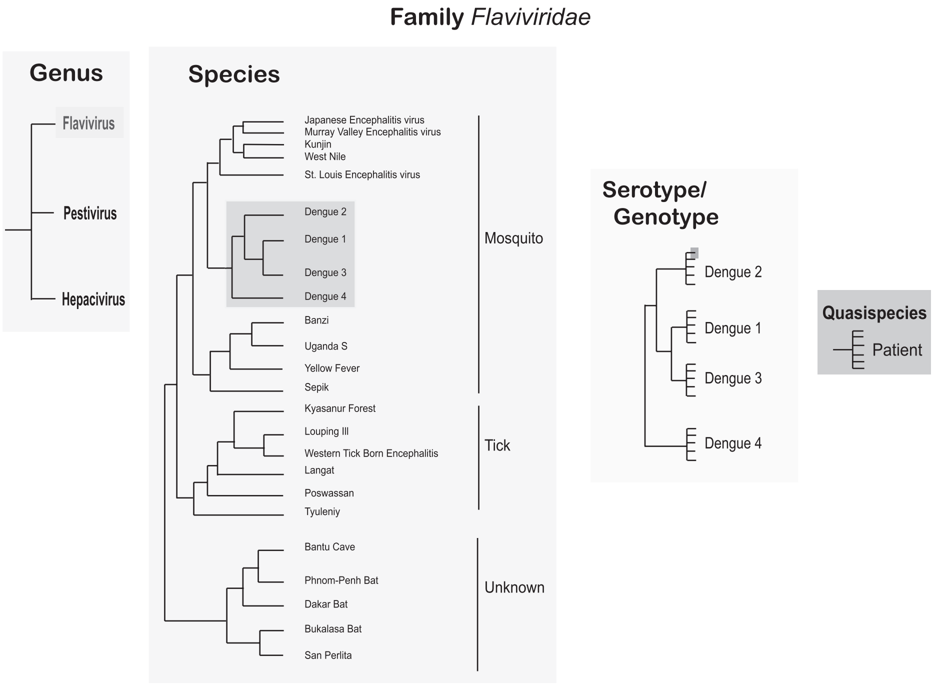
MOLECULAR EPIDEMIOLOGY

Traditional epidemiology relies on the combination of clinical presentation of disease, identification of the pathogen, and anecdotal circumstances to explain where the infection began and how it spread throughout a population. Common features are established to link the transmission to the source of infection. The addition of molecular tools to the investigative effort allows the infectious agent to be identified at the genetic level and enhances our understanding of virus origin, emergence, and transmission.

Molecular epidemiology uses phylogenetic methods to reconstruct a path of virus transmission based on heredity. As a virus replicates and moves through a population, mutations accumulate. The genetic variability displayed by the viruses is compared to deduce common ancestors and explain how one virus sequence gave rise to another. Viruses evolve, sometimes very quickly, to adapt to new hosts and environments. The genetic makeup of the pathogen may increase in variability with time and passage. Closely related virus sequences therefore correlate with recent infection and transmission. Using PCR and molecular tools, virus sequences are collected during an outbreak from infected human, animal, or insect reservoirs. These new virus sequences are then compared with a database of known virus sequences. The comparison is not a simple match of nucleotides between the virus genomes; rather, it involves sophisticated algorithms creating clusters of virus sequences sharing a common evolutionary ancestor.

Phylogenetic trees display how a set of virus sequences might have been derived during evolution, and provide guides in the placement and classification of an unknown virus. The trees are a graphical illustration of the evolutionary linkage of newly isolated virus sequences with known virus genomes (Figure 4.4).[10] The outer nodes of the tree display existing virus sequences. The inner branches represent theoretical ancestral virus sequences that gave rise to the recently isolated virus genome. The length of each branch corresponds to the amount of genetic change between the ancestral virus and the currently circulating virus. Additional algorithms may be applied to place a temporal scale with the amount of evolutionary change. A viral sequence that is very distantly related may be used as an outgroup to orient the tree with a direction of evolutionary change. Phylogenetic analyses revealing ancestry enable initial hypotheses about its basic life history, including the virus hosts and transmission patterns, as close relatives tend to be similar in their biology.

The phylogenetic tree can be interpreted as clusters of virus sequences that are used in classification schemes, and often these groupings complement traditional methods. Broad virus families can be defined using sequences from conserved genes such as polymerases and other enzymes required in the virus life cycle. Finer distinctions are noted by using sequences from genes that are more specialized for a particular virus. For example, envelopes and genes encoding structural proteins are often used to define virus subgroups within larger families. These subgroups often correspond to traditional serogroups defined by traditional serological methods that define groups by their ability to cross-react to a particular antibody. Entire genes or portions of genes can be used in the phylogenetic analysis, but complete virus genomes


Family *Flaviviridae*
Genus
Flavivirus
Pestivirus
Hepacivirus
Species
Japanese Encephalitis virus
Murray Valley Encephalitis virus
Kunjin
West Nile
St. Louis Encephalitis virus
Dengue 2
Dengue 1
Dengue 3
Dengue 4
Banzi
Uganda S
Yellow Fever
Sepik
Mosquito
Kyasanur Forest
Louping Ill
Western Tick Born Encephalitis
Langat
Poswassan
Tyuleniy
Tick
Bantu Cave
Phnom-Penh Bat
Dakar Bat
Bukalasa Bat
San Perlita
Unknown
Serotype/ Genotype
Dengue 2
Dengue 1
Dengue 3
Dengue 4
Quasispecies
Patient

Adapted from Holmes and Twiddy. Infection, Genetics and Evolution., 3:19-28. 2003

FIGURE 4.4 Phylogenetic tree of the RNA virus family *Flaviviridae*. Originally defined by serological assays, the phylogenetic tree shows inferred genealogical relationships based on maximum likelihood analysis of nucleic acid sequences. The family is divided into separate genera, *Pestivirus*, *Hepacivirus*, and *Flavivirus*. Flaviviruses are further subdivided into species that fall into three groups that can be transmitted via ticks or mosquitoes or without an arthropod vector (black vertical lines). Dengue viruses are separated into four serogroups or genotypes based on serology and envelope protein divergence. Within each Dengue serotype/genotype, representatives responsible for outbreaks of disease have been isolated from around the world. Given that RNA viruses mutate rapidly, a given isolate from an infected individual exists as a nonhomogeneous population variant termed a "quasispecies." Figure adapted from Ref. (10).

are rarely used, as the sequence length is restricted by the amount of computational time.

Epidemiological questions can be resolved with phylogenetic trees. The most fundamental questions in epidemiology are the mode of transmission and the origin of an outbreak. As the virus is passed through a population, genetic differences accumulate. More recent infections generally correspond to more shared derived genetic changes, and the most closely related will cluster on a tree. Transmission patterns are revealed as virus sequences from isolates are compared with sequences from a database and the clusters of sequences reveal a common ancestor.

An immediate question to solve during a virus outbreak is the mechanism by which the virus spreads. Viruses frequently infect animal or insect vectors that serve to pass the virus to humans. By identifying the virus through sequence analysis, a hypothetical reservoir can be predicted by the placement of the sequence on the phylogenetic tree, since viruses that share a mode of transmission often cluster together. More than one type of vector may be used within the same family of viruses, but individual members depend on a particular vector. Viruses from the family *Bunyaviridae* can be transmitted to humans by such pests as ticks, mosquitoes, flies, and rodents. The particular vector utilized often distinguishes individual subgroups within the family. Phylogenetic trees have been instrumental in proving virus transmission through family members, hospital settings, and susceptible members of a population. By comparing the virus sequence isolated from a patient to virus sequences isolated from individuals in the population transmission routes can be deduced. For example, it can be concluded that a doctor became infected from a patient seen at a hospital rather than from the general community if the virus sequences from doctor and patient are more similar than sequences found in the population.

The geographic location of virus infection can also be predicted by the use of phylogenetic trees. Viruses isolated from reservoirs in different geographic locations can define where people became infected. Arenaviruses are carried by rodents, but domestic rodents are often responsible for causing infections rather than rodents found in the fields or forests. By analyzing virus sequences obtained from infected patients and from house, field, and forest mice, the group of mice serving as a reservoir can be determined, and public health measures to prevent rodents from entering homes can be implemented. The origin of a virus outbreak can involve larger geographic areas. As the SARS outbreak demonstrates, importation of new viruses is an increasing global health concern. By accessing a database of virus genomes from around the world, a new virus outbreak could be rapidly identified as having been imported from another land. As viruses travel, some viruses become endemic, and it is necessary to distinguish between endemic and imported outbreaks to allow appropriate control measures to function. Viruses may also cause seasonal outbreaks. Influenza varies its genome with each year of infection. By understanding the extent to which the virus alters its genome from year to year, phylogenetic analysis assists in predicting the next year's strain and allows scientists to begin production for the upcoming yearly vaccine.

EXAMPLES OF NATURALLY EMERGING VIRUSES

SIN NOMBRE

In May 1993, a clinician reported a cluster of cases presenting with nonspecific illness, fever, headache, and cough that rapidly progressed to respiratory distress. The patients were young, previously healthy, and lived in rural areas in the four-corners region bordering Arizona, Colorado, New Mexico, and Utah. By early June 1993, 24 cases had been reported with a mortality rate approximately 75%.[11,12] Patient blood and autopsy tissues were sent to the Centers for Disease Control (CDC) for investigation. Diagnostic assays failed to identify known causes of respiratory illness; however, patient samples cross-reactived with antibodies to four different hantaviruses. Using hantavirus sera, immunohistochemistry detected hantavirus antigen in autopsy tissue from the lung and other involved organs.[13] Hantaviruses are RNA viruses with a genome in three segments and a member of the family *Bunyaviridae*. Previous outbreaks of hantaviruses involved patients with renal disease leading to hemorrhagic complications, not respiratory distress. Hantaviruses are classified into four antigenic groups, and each strain of hantavirus is tightly associated with a specific rodent host. Although hantaviruses have a worldwide distribution, the dependence on a particular species of rodent host restricts the geographical range. Hantaan (HTN) is associated with the field mouse in Korea, China, and eastern Russia and causes hemorrhagic fever with renal syndrome (HFRS). Seoul (SEO) causes moderate disease in Korea and China and is associated with urbanized rats. A milder form of HFRS is caused by the serogroup Puumula (PUU) in Scandinavia and Europe and is transmitted by the bank vole. At the time of the four-corners outbreak, the only North American hantavirus identified was Prospect Hill (PH), which was detected in two rodent isolates of meadow voles and not associated with any human illness. Patient samples from the U.S. four-corners region cross-reacted with hantavirus antigen and allowed epidemiologists to expand their assays to include RT-PCR and narrow the search for the host to local rodents. The virus was unable to be cultured from patient samples; therefore, molecular identification was crucial for identification. Partial sequences were available for HTN, SEO, PUU, and PH in a database and used to design PCR primers based on conserved areas of the virus glycoprotein. Nucleic acid was extracted from autopsy tissue, and a PCR product was detected in two cases using primer pairs based on PUU and PH. The PCR product was sequenced, revealing a 70% sequence similarity between the new virus and PH and PUU.[14] New primers were designed and could detect products in all current cases. The sequence of the new virus was identified, and the nucleocapsid protein was expressed in bacteria and purified and used as a recombinant antigen in other serological assays.[15] The new virus was eventually named Sin Nombre virus (SNV).

Rodents were trapped in areas in and surrounding patient houses. The predominant species was the field mouse, *Peromyscus maniculitis*, and approximately 30% had anti-hantavirus antibodies detected in an IgG ELISA. PCR revealed that 82% had detectable virus in the blood, indicating high viremia even in the presence of circulating antibodies. Sequences from rodent and human cases isolated in the four states were aligned, and identical nucleotide substitutions were present in the rodent and human cases from the same location, indicating that transmission occurred from the field mouse to the human cases.[14] Transmission occurred by inhalation of aerosolized rodent urine. Ecologists reported an increase in the population of field mice during that year due to a surplus food supply. The arid conditions of the southwest and the close proximity of the mice and the patients also contributed to transmission. In this outbreak, there were no reports of human-human transmission.[16]

The discovery of SNV led to a heightened awareness and search for other American hantaviruses. Field investigations and retrospective serological studies identified additional hantaviruses in North America.[17] The North American viruses show an 80%–95% similarity, but are always found only in their specific rodent species.[18] The detection of the new hantaviruses is enhanced by the use of diagnostic reagents developed in this outbreak and refined as new viruses are identified. Additional cases have also been detected in South America. In 1995, Andes virus was isolated in Argentina, the only documented case of human-human transmission.[19] Hantavirus outbreaks have also been identified in Chile, Paraguay, Bolivia, and Brazil, causing both respiratory illness and some renal hemorrhagic disease. South American cases show higher virus titers and also persistently infect a distinct rodent species as a natural reservoir.

After the SNV outbreak, efforts increased to study the relationships between hantaviruses and the rodent hosts worldwide. Phylogenetic analysis of rodent and human virus sequences has shown a close evolutionary relationship, and hantaviruses are often used as models of coevolution. A particular genotype of hantavirus persistently infects one species of rodent even though other rodent hosts may be present in the same environment. This remarkable coexistence has been mapped through changing landscapes. As a rodent species radiates farther away into another habitat, it may evolve into a separate, but genetically similar, species. The hitchhiker hantavirus also evolves with the rodent host, becoming genetically distinct from viruses associated with the ancestral species. This ancient relationship has been mapped using phylogenetic trees, which display common ancestral species as clusters that diverge and separate with geographical and genetic distance. The topology of the rodent tree can be compared and reveals similar patterns with topology of the associated hantaviruses.[20] Phylogenies have been performed on rodent family, subfamily, genus, and species using mitochondrial DNA. Phylogenetic trees of hantaviruses reveal a similar branching pattern to the rodent tree. Viruses that infect the same subfamily of rodents form identical clades, and closely related hantaviruses are grouped in the same tree pattern as their rodent reservoirs. Although rare, there have been instances of transmission of one virus to another rodent species. This provides another example of molecular evolution where virus segments can mix in the host and form reassorted viruses containing combinations of different virus segments. This can contribute to genetic diversity. Reassortment of the segmented genome has been found in nature and in mixed virus infections in tissue culture, but only between very closely related viruses.[21] Phylogenetic analysis performed on more than one virus segment tests for the chance of reassortment. Even in geographical locations where different rodent species may overlap, virus-host switching and reassortment are very rare events. Each hantavirus is associated with a primary rodent host, and usually a particular rodent species will be the primary reservoir of a single hantavirus. The tight correlation of virus and rodent host in the phylogenetic trees supports the concept of coevolution between virus and host. All subfamilies of the family *Muridae* have a hantavirus-rodent relationship. Fossil records estimate the rodent families *Muridae* to have split into the current subfamilies during the Miocene period, indicating that hantaviruses have been associated with their rodent hosts for 30 million years, an association that began before the division of the rodent family.[22] The coevolutionary relationship and worldwide distribution of hantavirus illustrates that the 1993 SNV outbreak was not due to the emergence of a new virus or virus with an increased virulence, but simply the detection of an ancient virus.

NIPAH

Between the autumn of 1998 and spring of 1999, an outbreak of acute fever with encephalitis occurred in peninsular Malaysia and Singapore, with 265 cases resulting in 105 deaths. The outbreak was associated with a respiratory illness of pigs, and the primary human cases were individuals who had close contact or occupational exposure to pigs.[23] The initial report was released in late September 1998, attributing the cause of illness to mosquito-borne Japanese encephalitis virus (JE). JE is endemic to the region and was believed to involve pigs as amplifiers of virus. Ministries of Health began mosquito control measures and vaccination campaigns to block JE infection.[24] By February 1999, the outbreak continued to spread as pigs were moved from the western-central region in the state of Perak to three other farms located in the southern states of Selangor and Negri Sembilan, the largest pig-rearing areas in Southeast Asia. Pigs were also exported to two farms in Singapore and resulted in 11 cases and two deaths. Singapore suspended pig importation from Malaysia, and no more cases were reported. Epidemiological analysis discredited JE as the infectious agent of the outbreak, since communities located near the pig farms did not become infected and immunization and vector control programs were having no effect on the increase of cases. As the outbreak continued, new attempts were made to find the pathogen.[25] In March 1999, a virus was isolated by inoculating CSF obtained from fatal cases into Vero cells. The presence of multinucleate syncytia in infected cell culture implicated a paramyxovirus. Thin-section EM detected filamentous nucleocapsids with the hallmark herringbone pattern and also captured virus budding from the plasma membrane, both characteristic of paramyxoviruses.[26] The new virus was named Nipah. A fortuitous event in the identification of the virus was the observed cross-reactivity between sera against Nipah and the paramyxovirus, Hendra virus. No other paramyxovirus sera reacted with Nipah samples.[25]

Hendra virus was identified in 1994 as the cause of respiratory illness in three regions of Queensland, Australia, resulting in the deaths of thoroughbred horses. A horse owner also died of respiratory distress and a stable-hand was infected, but survived. Another individual who assisted with a horse autopsy became ill with a meningitis-like infection and recovered (temporarily), but one year later developed progressive neurological complications and eventually died.[27] Hendra virus was isolated from his CSF. To test for other potential infections, reagents for an IgM and IgG ELISA and antibodies for IHC were developed.

Hendra-specific antibodies and antigen were used to screen initial Nipah cases. Eventually, Nipah-specific reagents were developed for larger serosurveys. In pig tissues, the major source of antigen and pathological damage was seen in the epithelium of the upper and lower airways. Most of the pigs in case farms were IgG-positive for Nipah, yet only a minority displayed any illness and only 5% were fatal.[28] Pigs had a definite respiratory disease comprising a distinctive cough with some developing encephalitis. The primary mode of transmission was suspected to be from respiratory secretions and aerosolized droplets from infected pigs. Domesticated dogs and cats, a few ponies, and bats also tested positive for Nipah. Human autopsy results revealed that the primary pathology was secondary to multiorgan vasculitis due to infected endothelium where multinucleate syncytia with viral antigen were detected. The cerebral cortex and brainstem were also damaged, and neurons were found to have virus inclusion bodies. Nipah antigen was also detected, but to a lesser extent in the lungs, heart, kidney, and spleen, consistent with a systemic infection.[26]

An RT-PCR assay was developed using degenerate paramyxovirus primers annealing to the P gene. Sequencing revealed a 78% similarity between Nipah and Hendra, but less than 50% identity to other members of the *Paramyxovirinae* subfamily.[29] The *Paramyxoviridae* family is divided into two subfamilies, the *Paramyxovirinae* and the *Pneumovirinae*. The *Paramyxovirinae* is separated into three genera; however, phylogenetic analysis did not cluster Nipah and Hendra into any one of the three genera. The *Paramyxovirinae* have a genome size around 15 kb, while both Hendra and Nipah have genomes of 18.2 kb due to a long untranslated region at the 3′ end of the genome.[30] The genome size and phylogenetic differences place Nipah and Hendra into the newly created *Henipah* genus.

The origin of the outbreak was traced to large fruit-eating bats, which are typically named flying foxes. Virus was isolated from urine collected under roosting areas and saliva from dropped fruit by the island flying fox, *Pteropus hypomelanus*. Antibodies have also been detected in the Malaysian flying fox, *Pteropus vampyrus*. Pig farms have fruit trees located near open pig stalls.[31,32] Flying foxes may transmit virus by urine droplets near the pig stalls or food source, or the pigs eat fruit contaminated with flying fox saliva. The outbreak was stopped after culling 1.1 million pigs.

WEST NILE

The West Nile virus outbreak illustrates several important points. First, it illustrates the difficulty in identifying a virus, especially one that had never been seen before in the U.S. Second, it shows that even with an initial misidentification, state and federal agencies were still able to act effectively, knowing how similar viruses could be controlled. And last, this case illustrates that virus distribution can be tracked by monitoring those infected as well as its vector.

In late August 1999, eight concurrent cases of patients having encephalitis and/or profound muscle weakness surfaced in Queens, New York. Geographically, the reported cases were traced to a two by two-mile area of a residential neighborhood, immediately sparking exposure concerns. By the end of the year, 59 patients were hospitalized in New York City, and seven deaths were reported.[33] Cases presented with a mild 3–6-day symptomatic period, including sudden onset of malaise, nausea, vomiting, headache, rash, cough, and sore throat. Less than 1% of those infected developed neurologic disease (i.e., encephalitis);[34] reported deaths were among the young, elderly, or immunocompromised. Also during this initial period it was noticed that several birds throughout the New York area were dying. The birds seemed to be dying from a neurological condition. Samples were sent to the CDC for identification.

Shortly after initial cases were reported, the CDC believed that the infectious agent was likely to be the endemic St. Louis encephalitis (SLE) virus. The CDC based their conclusion on serologic (IgM-capture ELISA) findings, which tested for the presence of patient antibodies to SLE antigens. The laboratory findings of SLE as the infectious agent were consistent with the symptoms of those infected, as well as the fact that SLE was responsible for 4,478 confirmed cases in the U.S. from 1964–1998.[35] SLE is an arbovirus, or a virus transmitted to human via arthropods, such as mosquitoes. Steps were quickly taken to monitor and control mosquito populations.

However, shortly after their initial assessment, the CDC changed their conclusion, and stated that the infectious agent was West Nile virus (WNV), not SLE. They based this on the more discriminating method of sequencing viral RNA isolates via RT-PCR, as well as results from specific monoclonal antibody detection of WNV-specific antigens (envelope glycoproteins). The controversy continued as other independent researchers believed the virus to be Kunjin virus, another closely related arbovirus.[36] It was later confirmed that the agent was indeed WNV through independent genetic testing by the U.S. Army Medical Research Institute for Infectious Diseases (USAMRIID).[33]

The cause for confusion was due to the close relationship between WNV and SLE. Both flavivirieses share significant similarity in their envelope proteins and some antibodies cross-react. Another compounding factor was that WNV had never been seen before in the U.S., although it was responsible for outbreaks in other geographic locations, including Africa, Europe, and the Middle East.

The diagnosis of WNV infection is made by both clinical findings and serologic tests, which detect patient IgM antibodies present in CSF or serum.[34,37] Direct virus isolation has been used, but due to limited sensitivities, is not a primary diagnostic procedure. RT-PCR methods for detecting WNV-specific RNA have been used extensively for detection in tissues and CSF. Blood banks now use RT-PCR to screen the blood supplies, due to the high incidence of asymptomatic WNV-infected donors: 23 persons have been reported to have been infected with WNV after receiving transfusions from 16 WNV-infected blood donors.[38,39]

West Nile virus was also identified as the agent responsible for the deaths of birds observed throughout the New York area. This has become a very important observation, for geographical mapping of the dead birds is valuable for tracking WNV infection as it continues to spread to other states and countries. The CDC stated: “dead-bird-based surveillance has proven to be the most sensitive method for detecting WNV presence in an area.” However the CDC goes on to state “mosquito-based surveillance remains the primary tool for quantifying the intensity of virus transmission in an area, and should remain the mainstay in most surveillance programs for WNV and other arboviruses.”

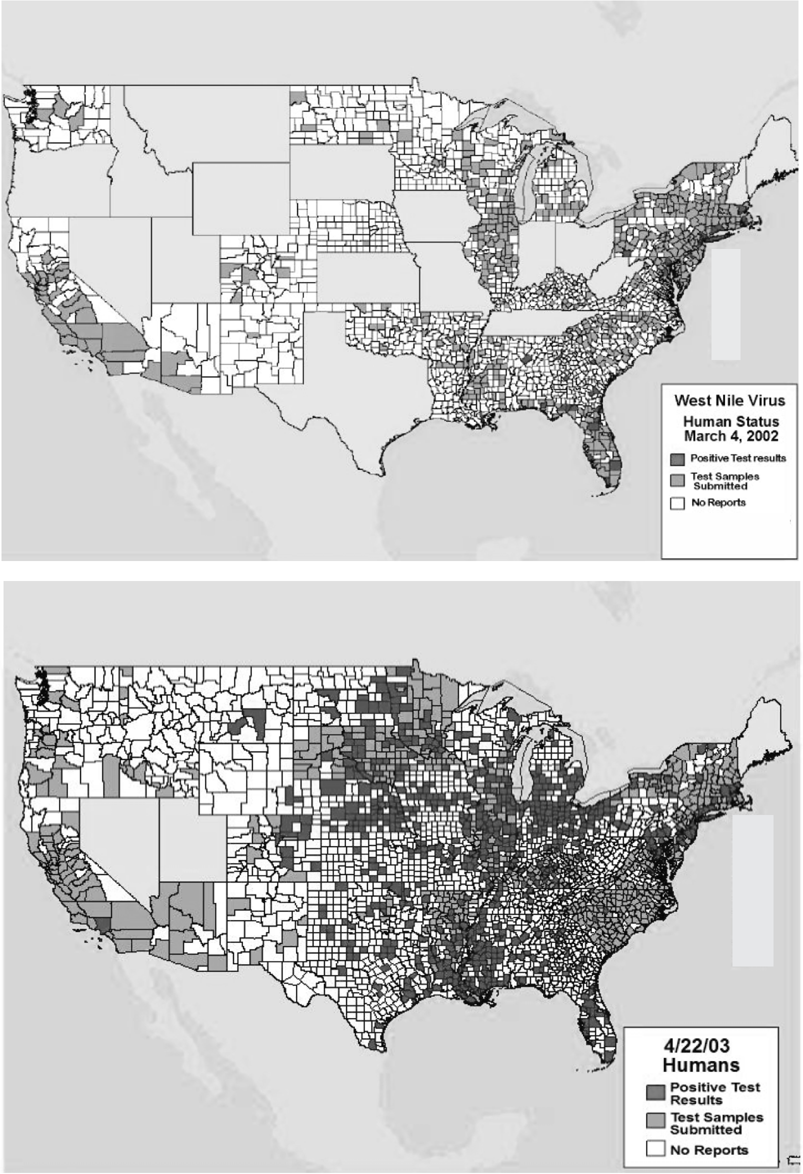

West Nile Virus
Human Status
March 4, 2002
Positive Test results
Test Samples Submitted
No Reports
4/22/03
Humans
Positive Test Results
Test Samples Submitted
No Reports

FIGURE 4.5 Spread of West Nile virus. From the initial outbreak in New York in 1999, the maps chart the rapid westward spread of WNV by showing the geographic distribution of human cases in years 2002–2004. Courtesy of the CDC. (See color insert.)

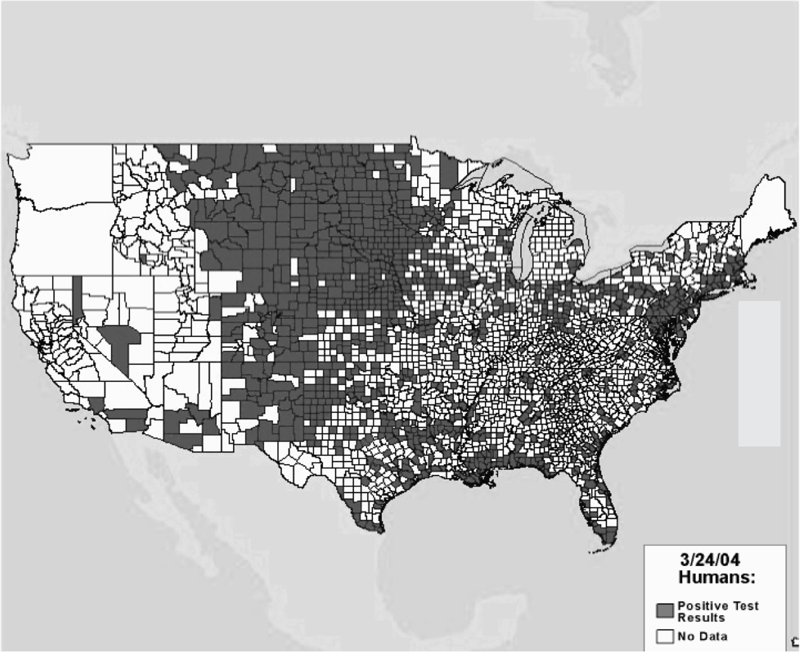


FIGURE 4.5 (*Continued*)

In addition to mosquito and dead-bird surveillance, the CDC as well as state organizations tracked live birds, equine WNV infections, and obvious human infections. These data showed that the virus was continuing to spread westward (Figure 4.5). Computer-assisted tracking services like Arbonet and the U.S. Department of the Interior website,[40] provide up-to-date tracking information. Such data allow for efficient knowledge acquisition regarding geographic distribution and changes in virus intensity, which gives the public the ability to take necessary precautions.

WNV continues to be a problem with over 4,000 cases reported in the U.S. in 2002. Accurate up-to-date tracking of infected humans, animals, and mosquito populations will continue to provide extremely valuable information about virus spread, locations at risk, and the efficiency of controlling arboviruses through mosquito control. Surveillance efforts must be supported to allow effective action against WNV and against future arbovirus epidemics as they occur.

SARS

The worldwide outbreak of SARS demonstrates the importance of rapid diagnosis of an infectious disease to understand how the disease spread and could be contained. The SARS outbreak was confounded by many problems. It is spread by close person-to-person contact allowing localized outbreaks in households and hospitals, and it has an incubation period of 7–10 days, allowing infected individuals enough time to travel anywhere in the world. The symptoms are very general: a high fever and cough or difficulty breathing, making it difficult to trace the spread by clinical presentation.[41] During the early course of the outbreak, investigators did not know the agent causing the infection.

In November 2002, several hundred cases of severe, atypical pneumonia were reported from the Guangdong province in southern China. An infected medical doctor from Guangdong carried the virus to Hong Kong, where he stayed for one night on the 9th floor of the Hotel Metropole. Subsequently, at least 16 guests and visitors to the 9th floor of the hotel became infected and spread the illness to Vietnam, Singapore, and Canada. Hospitals became new epicenters of infection as health care workers and family members became infected. In March 2003, the World Health Organization (WHO) recognized that SARS had become a worldwide health threat, and released a global alert on travel to screen air passengers departing from affected areas. Cases of suspected SARS continued to be reported in new areas including Thailand, Europe, the United Kingdom, and the U.S., but the most heavily affected areas were in China and Hong Kong. In early April, the virus was isolated and identified as a novel coronavirus, SARS-associated coronavirus, or SARS-CoV. The numbers of cases increased at a staggering pace with over 5,000 cases on April 28, 6,000 cases on May 2, and 7,000 cases on May 8 from over 30 countries on six continents.[42] Travel warnings were issued for Toronto, Canada, where an imported case of SARS established an outbreak in a hospital with over 100 cases, which subsequently spilled over into the community. Travel restrictions also applied to Beijing and Guangdong, China and Hong Kong. Although transmission was assumed to be primarily from secreted virus in respiratory fluids, another mode of transmission was found in an outbreak in a Hong Kong apartment building. Improper plumbing in the Amoy Gardens apartment complex probably allowed virus-contaminated sewage to be introduced into the water supply, resulting in almost 300 cases from one building. Epidemiologists began to link a few individuals as sources responsible for infecting a large number of people. These “super spreaders” are individuals who for unknown reasons were capable of transmitting infection to over 100 people. In Singapore, one individual traveling from the Hotel Metropole was credited for infecting 103 out of 203 cases. Despite these few instances, transmission was primarily due to close contact with infected individuals and remained in the hospital epicenters where import cases had harbored the virus. Rarely, secondary outbreaks in the community occurred. This required containment measures to be enacted. Local authorities began voluntary and mandatory quarantine procedures and eventually, the number of new cases began declining. Officials declared regions to be free of SARS when no new cases were identified after 20 days, twice the time of an incubation period. By early July 2003, there were no new cases of SARS reported. As of September 2003, WHO reported an estimated 8,098 cases and 774 deaths from the SARS-CoV.[43]

The rapid identification of a virus as the etiologic agent of SARS allowed public health officials to address the public on how to kill the virus, any treatment options, and use knowledge of other animal coronaviruses to form hypotheses regarding origin and transmission. The virus was isolated by inoculation of Vero E6 and NCI-H292 cells from an oropharyngeal specimen obtained from a fatally infected SARS patient.[44] The cells showed CPE with cell rounding in focal areas that were refractile to light. The infected cell cultures were treated for thin-section and negative-stain electron microscopy, where the particles were detected with the distinctive halo of the *Coronaviridae* (Figure 4.3). Investigators were extremely fortunate to have an isolate propagating and causing an obvious cytopathic effect in cell culture. Coronaviruses are notorious for being difficult to culture, requiring specific cell lines with supplements, organ culture, or animals. The characteristic particles identified using electron microscopy enabled researchers to swiftly focus on a specific virus group. *Coronaviridae* have three serological groups. Using infected Vero cells as a source of SARS-CoV antigen, antibodies specific for each coronavirus group were tested. Only antibodies from group I recognized SARS-CoV-infected cells. Antibodies from groups I, II, and III all failed to detect virus antigen in autopsy samples despite the presence of syncytial cells and other signs of pathology. Infected Vero cells were again used as antigen to detect the presence of SARS-CoV antibodies from patient samples. Patient sera with high titers of SARS-CoV-specific antibodies were used to develop an ELISA. SARS-CoV antibodies were later raised in animals injected with virus infected cell culture. Random samples from the population did not have detectable IgG antibodies, implying that SARS-CoV had not previously been circulating and instead was a newly emerged virus.[44]

The SARS-CoV was also identified using RT-PCR with a broadly reactive primer set designed based on the polymerase gene. The entire genome was sequenced and phylogenetically compared to other coronaviruses. SARS-CoV was placed equidistant from all other animal and human coronaviruses in all gene regions. This implied that the SARS-CoV did not arise from recombination from any known coronavirus, nor did it help identify the host. To verify

SARS-CoV as the actual agent responsible for disease, two cynomolgus macaques were inoculated with a Vero cell culture infected with serum from a fatal SARS case. SARS-CoV was detected in nasal secretions and fecal material by RT-PCR and immunofluorescence. The virus sequence was identical to the inoculated virus.[45] Histopathology showed cell damage in the lungs, and both macaques seroconverted, proving Koch's postulate for SARS-CoV as the causative agent.[46]

With SARS worldwide, the number of samples requiring diagnostics grew. Designation of SARS-CoV as a biosafety level 3 virus required specialized laboratories, so the WHO coordinated diagnostic labs from nine countries to work together to standardize diagnostic assays, supply similar reagents, ensure proper equipment availability, and analyze results in a comparable manner. Communication was vital with daily teleconferences, secure web sites, and exchange of patient samples and autopsy materials.[43]

The SARS epidemic demonstrates how a delay in reporting and an incorrect diagnosis can allow an epidemic to spread. Initially, the Chinese government misdiagnosed the infectious agent as Chlamydia and underestimated the ease of transmission and potential dispersal throughout the region. Epidemiologists from the WHO were not permitted into the Guangdong province until May 2003, but by then the virus had spread all over the globe. Investigators were distracted by reports of fatalities from a virulent strain of avian influenza, which had resurfaced in Hong Kong in late February 2003. Also, the human metapneumovirus, which also causes respiratory illness, had been identified in cell cultures inoculated with some samples from SARS patients. It was later concluded that the newly identified coronavirus was responsible for SARS, and the metapneumovirus was simply circulating concurrently in some populations.

It is important to find the natural reservoir of the SARS-CoV to prevent its reintroduction into the human population. Retrospective epidemiology of the initial outbreak in Guangdong, China revealed that early cases were local restaurant workers handling wild animals. Investigators tested animals from local marketplaces and performed a serological survey of animal handlers. The SARS-CoV was isolated from the Himalayan palm civet, and antibodies were detected in a raccoon-dog and a Chinese ferret badger as well as some animal handlers. The sequence from the palm civet contained a 29-nucleotide region that was not present in the genomes of virus isolated from the SARS epidemic.[47] It is unknown whether this region contributed to the transmission to humans. It cannot be assumed that the palm civet or other animals in the marketplace are the actual natural reservoir or just infected bystanders.

The future of SARS is uncertain. We still do not understand the natural history of the SARS-CoV. We cannot predict whether and when the virus will reemerge. The virus may cause seasonal outbreaks or establish endemic disease in rural health care settings. Researchers are assured the virus will appear again and are working to find therapies and a vaccine.

MONKEYPOX

The following case illustrates how trace-back investigations can determine how an African virus infected 81 people across the Midwest in the U.S. On May 20, 2003, in Wisconsin, a young girl was taken to the hospital with a rash and raised fluid-filled bumps on her hand that formed after a bite from her pet (Figure 4.6). The doctor who treated her suspected that the girl may have been infected with either the plague or tularemia. Shortly thereafter, both the mother and father became infected with the unknown agent. On May 30, 2003, upon viewing EM micrographs of the mother's biopsy as well as cell culture, state doctors believed the infectious agent was a member of the orthopoxvirus genus. The EM micrographs (Figure 4.6) were useful in identifying the genus of virus, but were also of limited use for species identification because all members of the orthopoxvirus genus have similar morphology. The agent could have been any member such as cowpox, vaccinia, monkeypox, or even smallpox. On June 6, 2003, local clinicians were able to culture the orthopox virus obtained from the pet prairie dog's lymph nodes. Samples were quickly sent to the CDC in Atlanta, Georgia for identification.

While samples were awaiting identification, several more cases were reported. These were predominantly from Midwestern states, including

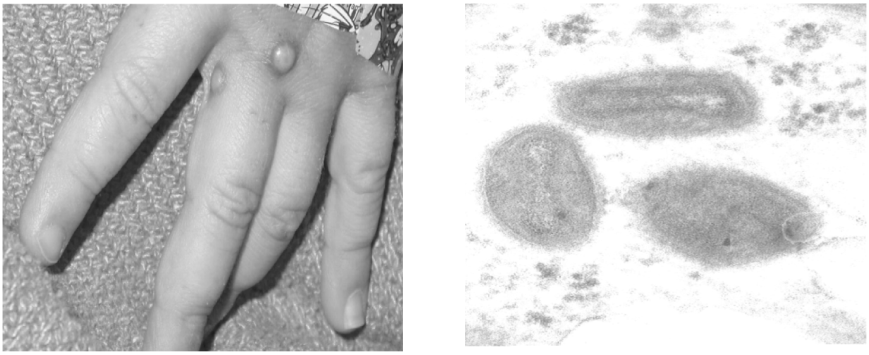


FIGURE 4.6 (A) Monkeypox skin lesion on a young girl's right hand. (B) Electron micrograph of orthopox virus cultured from her mother. Courtesy of the Marshfield Clinic, Marshfield, Wisconsin.

Wisconsin, Illinois, Indiana, Missouri, Ohio, and Kansas. Patients showed signs of rash, fever, cough, and headache, among other symptoms. As the patients were interviewed and information gathered, a common feature became apparent. Most patients had contact with exotic pet prairie dogs. On June 8, 2003, the CDC positively identified the virus as monkeypox.

By July 2, 2003, a final total of 81 monkeypox cases were reported, spanning several states. The CDC confirmed cases using a variety of laboratory techniques including propagation in tissue culture, PCR, immunohistochemical testing, and/or EM.[48] Monkeypox virus got its name when it was first discovered in 1958, after several laboratory monkeys became infected with the virus. There have been few reports of this virus, which is native to rainforest regions of Central and Western Africa. However, in 1996–1997 an outbreak occurred in the Democratic Republic of Congo (DRC), where monkeypox was positively identified in 72 cases spanning thirteen villages.[49] The virus has been found to infect native squirrels, shrews, monkeys, and Gambian rats among other animals representing 14 species.[48] The cause of the epidemic is a result of handling or consuming contaminated bush meat and foodstuffs. The virus was spread most likely through person-to-person contact, due to close-quarters habitation. Of the cases documented, six deaths occurred, accounting for a 3% case-fatality rate.[49]

In the U.S. outbreak several states linked the purchases of the prairie dogs to particular pet stores. The trace-back investigation revealed all of these pet stores purchased prairie dogs from a distributor in Illinois. Through PCR-based genetic testing, the CDC uncovered prairie dogs at the Illinois dealer tested positive for monkeypox. How did these prairie dogs become infected with a virus native to Africa? This distributor, on April 21, 2003, received a shipment of African Gambian rats and dormice. These were then tested by the CDC, and found to be infected with the virus. Investigators then were able to track the rodent shipment to a Texas distributor, which was the initial point of entry for several rodent species imported from Ghana on April 9, 2003. Several of these imported animals were infected with monkeypox. Finally a source for the monkeypox entry into the U.S. was located. A schematic illustration of the trace-back investigation is shown in Figure 4.7.

Very quickly case investigators reconstructed the events of the outbreak. Their trace-back investigation concluded that the method of virus spread to human populations was through contact with monkeypox-infected prairie dogs (animal's body fluids or skin rash); this was in contrast to the outbreak in the DRC where monkeypox spread predominantly via person-to-person contact. The prairie dogs were infected by the monkeypox-carrying rodents at the Illinois animal distributor. Because pet stores from many different states purchased from one distributor, the geographic region that became susceptible drastically expanded.

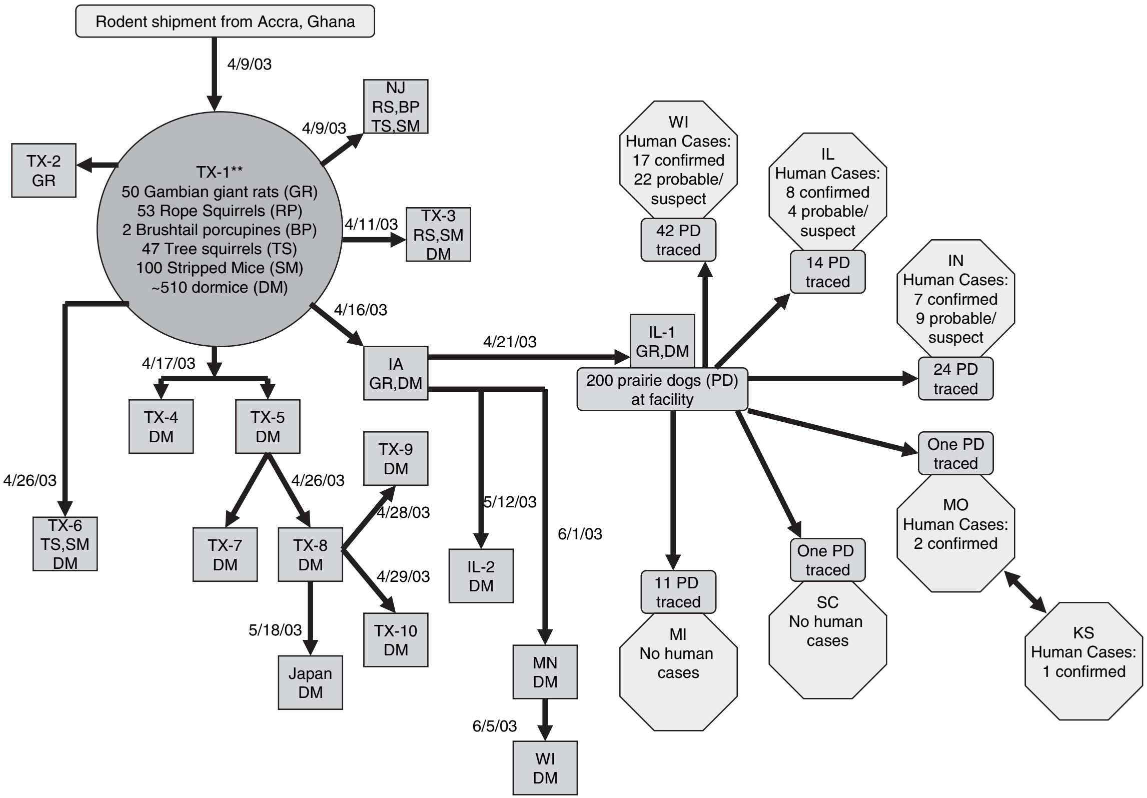

Rodent shipment from Accra, Ghana
4/9/03
TX-1**
50 Gambian giant rats (GR)
53 Rope Squirrels (RP)
2 Brushtail porcupines (BP)
47 Tree squirrels (TS)
100 Stripped Mice (SM)
~510 dormice (DM)
TX-2
GR
4/9/03
NJ
RS,BP
TS,SM
4/11/03
TX-3
RS,SM
DM
4/16/03
IA
GR,DM
4/17/03
TX-4
DM
TX-5
DM
4/26/03
TX-6
TS,SM
DM
TX-7
DM
4/26/03
TX-8
DM
4/28/03
TX-9
DM
4/29/03
TX-10
DM
5/18/03
Japan
DM
5/12/03
IL-2
DM
6/1/03
MN
DM
6/5/03
WI
DM
4/21/03
IL-1
GR,DM
200 prairie dogs (PD)
at facility
WI
Human Cases:
17 confirmed
22 probable/
suspect
42 PD
traced
IL
Human Cases:
8 confirmed
4 probable/
suspect
14 PD
traced
IN
Human Cases:
7 confirmed
9 probable/
suspect
24 PD
traced
One PD
traced
MO
Human Cases:
2 confirmed
KS
Human Cases:
1 confirmed
One PD
traced
SC
No human
cases
11 PD
traced
MI
No human
cases

FIGURE 4.7 CDC's trace-back investigation of the monkeypox outbreak. Adapted from the CDC.

VIRAL FORENSICS

THE SCHMIDT CASE

The value of phylogenetic analyses in forensics is well illustrated by the recent and curious criminal case of the State of Louisiana versus Richard J. Schmidt.[50] The uncontested facts in this case are that a gastroenterologist from Lafayette, Louisiana broke into the home of his former mistress and office nurse late at night on August 4, 1994, and that he argued with her and gave her an intramuscular injection. He claimed it was a vitamin B shot. She claimed it was HIV. She had begun feeling ill several months after the injection, and a blood test in January 1995 revealed that she had become infected with HIV. She was a periodic blood donor, and based on tests of those previous blood donations, had a clear record without prior infection. She did not engage in any behaviors placing her at high risk for infection, and her sexual contacts over the previous nine years all tested negative for HIV. He was a community leader and Vietnam War veteran. Schmidt admitted to having a long-term affair with her, but maintained the infection was not his doing.

She went to the District Attorney's office to file charges on learning that she was HIV-positive. Moving quickly, the DA's detectives obtained a search warrant and proceeded to the accused physician's office, where they seized his record books for blood samples drawn from patients and a vial of blood sitting in the refrigerator in a back room in his office. This was unusual; patients' blood samples were sent to the lab soon after being drawn, and none were routinely stored there. The physician claimed that this sample, drawn from one of his HIV-positive patients, was for his own use and research. Was the physician telling the truth? Might this blood sample link the physician to his nurse assistant and former mistress' infection? The next logical step in the investigation would clearly be phylogenetic analyses. Phylogenetic analyses of viral DNAs showing little or no relationship between HIV lineages from the nurse and the alleged source (the blood vial seized from the physicians office) could help demonstrate the physician's innocence, whereas a close, sister relationship among those lineages, in the context of an epidemiological sampling of HIV, would be consistent with the physician's alleged role in transmission.

Are phylogenetic analyses better than other, more routine methods of forensic analyses? In some instances, yes. Analyses based on similarity alone, such as comparison of genetic fingerprinting data in which restriction fragment patterns for hypervariable DNA sequences are compared, are subject to greater bias from similarity due to chance, known as homoplasy, rather than similarity due to common ancestry, known as homology. By contrast, phylogenetic analyses attempt, explicitly, to show the pattern of common descent among samples analyzed, rather than simple similarity. Phylogenetic analyses have the potential to show homoplasious similarity for what it is, convergence, and not to be misled, when the desired information is evolutionary relationships.

Portions of the HIV env gp120 gene and the RT genes were specifically chosen to be sequenced in an attempt to maximize the phylogenetic information in the dataset. The env gene evolves relatively quickly, being selected upon by hosts' immune systems, and is capable of recovering relationships among recently diverged HIV lineages. The RT gene evolves more slowly, due to greater functional constraints, and can provide insight into relatively older divergences. Using both sequences thus provides a broader range of evolutionary rates than either does alone.

Initial phylogenetic analyses were conducted using maximum parsimony with the software program PAUP,[51] and final analyses were conducted using an explicit model for sequence evolution seeking to account for heterogeneity in rates of change across nucleotide sites and across virus lineages. This was done using the maximum-likelihood optimality criteria in a bayesian context with the program MrBayes.[52] All analyses of the HIV env gp120 sequences from the nurse, the alleged source, and the epidemiological sampling of HIV patients were congruent in showing the HIV sequences from the victim to form a single monophyletic clade, in showing the alleged source to form a single monophyletic clade, and in showing those two clades to be closest relatives (sister taxa) relative to the epidemiological sample. This is consistent with the accusation that the physician used the blood sample from one of his patients to infect the nurse, although this rapidly evolving sequence provides no information regarding the direction of infection.

The more slowly evolving RT sequences also indicate their close evolutionary relationship, but with an additional and valuable piece of information. Based on RT sequences, viruses from the victim arose from within the clade of viruses from the alleged source. That is, the alleged source viruses are paraphyletic (incomplete in this case) unless the victim's viruses are included and nested within that group. This analyses does provide more direct evidence about the direction of infection, with the clear implication that viruses from the alleged source were used to infect the victim. Viral lineages from the alleged source diverged prior to divergences among the victim's viruses. This difference from the tree topologies based on env sequences can be traced to the slower rate of RT sequence change, with longer associated coalescence times for gene lineages, showing an earlier set of divergence events, as expected.

The phylogenetic analyses and rationale above are mainstream methods among evolutionary biologists; however, phylogenetic thinking and explicit use of evolutionary trees to track genealogy and transmission of virus or bacterial strains between individuals is not yet common in the U.S. courts. The case of the State of Louisiana versus Richard J. Schmidt set a precedent in this area. A pre-trial admissibility hearing was held regarding the proposed use of phylogenetic analyses in the criminal trial accusing Schmidt of attempted murder. Despite the efforts of the defense to block their admissibility, based on arguments that the viruses were evolving too rapidly to allow tracing of their shared ancestry (they were not), the judge ruled that phylogenetic analyses did meet judicial standards of admissibility, being subject to empirical testing, published in peer reviewed sources, and generally accepted within the scientific community.

Though not a panacea, phylogenetic analyses will prove useful in a range of forensic investigations. As with other molecular forensic approaches they can be particularly effective in demonstrating the innocence of accused individuals. They can be also be useful in tracing sources for any transfer of infectious materials whether viral, bacterial, or protozoan, involving accidental contamination or intentional infection in personal crimes or acts of terrorism. However, these applications are potentially limited by rates of sequence change, which must be sufficiently fast to provide a record of phylogenetic relatedness, but slow enough to preserve sufficient phylogenetic signal, prior to its being overwritten with multiple substitutions at individual sites. Application of phylogenetic analyses can also be complicated by the propensity of some viruses to recombine.

The defendant was found guilty of attempted murder and sentenced to 50 years in prison, the maximum allowable under the law. In this particular case, tried by jury, the phylogenetic evidence was consistent with the prosecution's case; however, there was other evidence that the jury may have found even more compelling. This included the physician having hidden the notebooks of his blood sampling and having a history of threats against the victim as she tried to end their affair. On March 4, 2002 the U.S. Supreme Court rejected an appeal of the verdict, thus establishing precedent for use of phylogenetic analyses in U.S. courts of law.

ENGINEERING NOVEL VIRUSES

BEFORE AND AFTER RECOMBINANT DNA

As mentioned earlier, viruses in nature, depending upon their replication strategy and host range, have many ways to evolve through mutation, recombination, reassortment, and selection. These innate properties have been used as tools in virology for decades. Examples include cold-adapted and attenuated live vaccine strains of influenza, cross-species rotavirus reassortants as vaccine candidates, and the attenuated yellow fever virus 17D vaccine strain that was derived by serial passaging in cell culture. Over the last 30 years, the recombinant DNA revolution has changed virology forever. Besides its impact on our ability to study specific viral nucleic acids and proteins, recombinant DNA technology (and PCR) sped up our ability to determine complete viral genomic sequences and create specific viral variants. The ability of scientists to engineer novel viruses varies greatly. Small positive-strand RNA viruses of bacteria, plants, and animals provided some of the first examples where infectious virus was produced via recombinant DNA. More recently, it has also become possible to engineer the >29-kb RNA genome of coronaviruses and both monopartite and segmented negative-strand RNA viruses. Viruses in the family *Reoviridae*, with double-stranded RNA genomes, have proven more difficult. For herpesviruses, whose genomic DNA is infectious, the same technologies used for cloning large pieces of chromosomal DNA have been applied successfully to propagate overlapping fragments of entire herpesvirus genomes. For the poxviruses, whose genomic DNA is not infectious, recombinant viruses are generated within infected cells by homologous DNA recombination between a plasmid containing the engineered segment and an infecting parental virus.

SYNTHETIC POLIOVIRUS

Typically, recombinant DNA manipulations of viruses begin with viral nucleic acid (RNA or DNA) that is then cloned and amplified in bacteria using plasmid vectors or *in vitro* using PCR. Once the genome sequence of a virus is known, however, it becomes possible to create this sequence artificially using overlapping synthetic oligonucleotides and gene synthesis techniques that have been available for many years. An example of this, which received a great deal of attention from the media and the scientific community, was recently published for an attenuated form of poliovirus.[53] While the poliovirus case came as no surprise to virologists and molecular biologists, it did bring several issues to the forefront. If an infectious virus can be created by synthetic methods, can it ever really be eradicated? In light of potential bioterrorist or biowarfare agents, should there be a restriction on making the genome sequence available in public databases? Should research on these agents be banned, classified, or otherwise regulated? These and other questions will continue to be debated in the years to come.

MOUSEPOX

The ability to engineer viruses raises the concern about modifying existing viruses to make them more virulent. The bulk of examples in virology run counter to this idea. Viruses are usually highly adapted to a particular niche, and most mutations are deleterious. Propagation of pathogenic viruses in cell culture often leads to adaptation to that environment and attenuation in their animal host. Recombinants or chimeras between even closely related viruses are usually impaired relative to either parent. We do see examples, however, of viruses that are benign in one animal host but highly pathogenic in another species. This is often the case in epizootic emerging viruses. An extreme example of host species-specific pathogenesis is the myxoma poxvirus, which causes a benign cutaneous fibroma in wild rabbits of the Americas but a highly lethal disease in the European rabbit. This virus was used for biological control of feral European rabbits in Europe and Australia. In 2001, an Australian group published a paper describing the construction of recombinant mousepox virus expression interleukin-4 (IL-4) and its pathogenesis in mice.[54] This study provides a striking and sobering example of an engineered virus with enhanced pathogenicity. IL-4 is a cytokine that regulates the immune response at various levels. The recombinant mousepox virus suppressed both innate and adaptive immune responses that normally control infection and was lethal for otherwise genetically resistant mice. Moreover, even animals that had been previously vaccinated and protected from virulent mousepox were susceptible to lethal infection by the mousepox-IL-4 recombinant virus. This study raises obvious concerns about the efficacy of current vaccination against modified versions of the smallpox virus.

DETERMINING THE SOURCE OF AN ENGINEERED VIRUS

Viral replication, mutation, and evolution both help and hinder determination of the origin of a virus. In the case of naturally occurring emerging viruses, sequence comparisons of isolates linked temporally to disease, geography, and species can help pin down the virus origin. Cases involving deliberate dissemination of a naturally occurring isolate, as in the Schmidt case, require not just phylogenetic comparisons but also other forensic evidence. The ability to engineer viruses makes the situation even more complex. While there are numerous laboratory strains of different viruses for which we know the exact sequence, the sequence of an isolate does not prove that it originated from a particular source. Viruses can be transferred and propagated or, as noted above, functional viral genomes can be created synthetically to mimic (or diverge from) a publicly available sequence.

CONCLUSION

In many ways, virology and viral forensics are still in their infancy. We know very little about the vast spectrum of viruses in nature. Although we should continue to pursue studies on known human pathogens, we need to broaden our efforts in virus isolation, viral genomics, and bioinformatics. A comprehensive database of viral sequences is critical for rapid identification of emerging viral pathogens and new diagnostic platforms, such as oligonucleotide arrays.[55] In addition to viral genomics, global proteomic analysis of viral disease processes may uncover molecular signatures that can be used to implicate a particular etiologic agent or class of agents, even in the absence of viral nucleic acid or serology. The global eradication of smallpox ranks as one of humankind's greatest accomplishments. In the wake of this triumph, many viral challenges remain. HIV, hepatitis B and C, influenza, dengue, rotavirus, and many other viral diseases continue to affect tens of millions. It is distressing that in the face of these undeniable challenges we are back to discussing the deliberate use of existing or modified viruses for harm, not good.

ACKNOWLEDGMENTS

Work by D.P.M. was supported by the National Science Foundation (DBI-9974525); J.H., L.K.M, H.L.H., and C.M.R. are supported by grants from the Public Health Service and the Greenberg Medical Research Institute.

REFERENCES

1. Baltimore D. Expression of animal virus genomes. *Bacteriol Rev* 1971;35:235–241.
2. Mayo MA, Pringle CR. Virus taxonomy–1997. *J Gen Virol* 1998;79 (Pt 4):649–657.
3. MHV van Regenmortel, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR, Wickner RB. Virus Taxonomy: The Classification and Nomenclature of Viruses. The Seventh Report of the International Committee on Taxonomy of Viruses. Vol. VIIth report of the ICTV. San Diego: Academic Press, 2000.
4. Domingo E, Holland JJ. RNA virus mutations and fitness for survival. *Annu Rev Microbiol* 1997;51:151–178.
5. Nathanson N, Fine P. Virology. Poliomyelitis eradication—a dangerous endgame. *Science* 2002;296(5566):269–270.
6. Kew OM, Nottay BK, Hatch MH, et al. Multiple genetic changes can occur in the oral poliovaccines upon replication in humans. *J Gen Virol* 1981;56(Pt 2):337–347.
7. Kew O, Morris-Glasgow V, Landaverde M, et al. Outbreak of poliomyelitis in Hispaniola associated with circulating type 1 vaccine-derived poliovirus. *Science* 2002;296(5566):356–359.
8. Morse SS. *Emerging Viruses*. New York: Oxford University Press, 1993.
9. Webster RG, Laver WG, Air GM, Schild GC. Molecular mechanisms of variation in influenza viruses. *Nature* 1982;296(5853):115–121.

10. Holmes EC, Twiddy SS. The origin, emergence and evolutionary genetics of dengue virus. *Infect Genet Evol* 2003;3(1):19–28.
11. Outbreak of Acute Illness-Southwestern United States, 1993. *Morb Mortal Wkly Rep* 1993;42(22):421–424.
12. Khan AS, Ksiazek TG, Peters CJ. Hantavirus pulmonary syndrome. *Lancet* 1996; 347(9003):739–741.
13. Zaki SR, Greer PW, Coffield LM, et al. Hantavirus pulmonary syndrome. Pathogenesis of an emerging infectious disease. *Am J Pathol* 1995;146(3):552–579.
14. Nichol ST, Spiropoulou CF, Morzunov S, et al. Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. *Science* 1993;262(5135):914–917.
15. Feldmann H, Sanchez A, Morzunov S, et al. Utilization of autopsy RNA for the synthesis of the nucleocapsid antigen of a newly recognized virus associated with hantavirus pulmonary syndrome. *Virus Res* 1993;30(3):351–367.
16. Mills JN, Yates TL, Ksiazek TG, et al. Long-term studies of hantavirus reservoir populations in the southwestern United States: rationale, potential, and methods. *Emerg Infect Dis* 1999;5(1):95–101.
17. Peters CJ. Hantavirus Pulmonary Syndrome in the Americas. *In* Scheld WM, Craig WA, and Hughes JM, eds. *Emerging Infections* 2. Washington, D.C.: ASM Press, 1998:17–64.
18. Morzunov SP, Rowe JE, Ksiazek TG, et al. Genetic analysis of the diversity and origin of hantaviruses in *Peromyscus leucopus* mice in North America. *J Virol* 1998;72(1):57–64.
19. Wells RM, Sosa Estani S, Yadon ZE, et al. An unusual hantavirus outbreak in southern Argentina: Person-to-person transmission? Hantavirus Pulmonary Syndrome Study Group for Patagonia. *Emerg Infect Dis* 1997;3(2):171–174.
20. Plyusnin A, Morzunov SP. Virus evolution and genetic diversity of hantaviruses and their rodent hosts. *Curr Top Microbiol Immunol* 2001;256:47–75.
21. Rodriguez LL, Owens JH, Peters CJ, Nichol ST. Genetic reassortment among viruses causing hantavirus pulmonary syndrome. *Virology* 1998;242(1):99–106.
22. Nichol ST. Genetic analysis of hantavirus and their host relationships. *In* Saluzzo JF, Dodet B, ed. *Emergence and Control of Rodent-borne Viral Diseases*. Annecy, France: Elsevier, 1998: 99–109.
23. Goh KJ, Tan CT, Chew NK, et al. Clinical features of Nipah virus encephalitis among pig farmers in Malaysia. *N Engl J Med* 2000;342(17):1229–1235.
24. Chua KB, Goh KJ, Wong KT, et al. Fatal encephalitis due to Nipah virus among pig-farmers in Malaysia. *Lancet* 1999;354(9186):1257–1259.
25. Chua KB, Bellini WJ, Rota PA, et al. Nipah virus: A recently emergent deadly paramyxovirus. *Science* 2000;288(5470):1432–1435.
26. Goldsmith CS, Whistler T, Rollin PE, et al. Elucidation of Nipah virus morphogenesis and replication using ultrastructural and molecular approaches. *Virus Res* 2003;92(1):89–98.
27. Field H, Young P, Yob JM, et al. The natural history of Hendra and Nipah viruses. *Microbes Infect* 2001;3(4):307–314.
28. Daniels P, Ksiazek T, Eaton BT. Laboratory diagnosis of Nipah and Hendra virus infections. *Microbes Infect* 2001;3(4):289–295.
29. Harcourt BH, Tamin A, Ksiazek TG, et al. Molecular characterization of Nipah virus, a newly emergent paramyxovirus. *Virology* 2000;271(2):334–349.
30. Harcourt BH, Tamin A, Halpin K, et al. Molecular characterization of the polymerase gene and genomic termini of Nipah virus. *Virology* 2001;287(1):192–201.
31. Chua KB, Koh CL, Hooi PS, et al. Isolation of Nipah virus from Malaysian Island flying-foxes. *Microbes Infect* 2002;4(2):145–151.
32. Yob JM, Field H, Rashdi AM, et al. Nipah virus infection in bats (order Chiroptera) in peninsular Malaysia. *Emerg Infect Dis* 2001;7(3):439–441.

33. Outbreak of West Nile-like viral encephalitis–New York, 1999. *MMWR Morb Mortal Wkly Rep* 1999;48(38):845–849.
34. Petersen LR, Marfin AA, Gubler DJ. West Nile virus. *JAMA* 2003;290(4):524–528.
35. http://www.cdc.govncidod/dvbid/arbor/sle_qa.htm
36. Briese T, Jia XY, Huang C, et al. Identification of a Kunjin/West Nile-like flavivirus in brains of patients with New York encephalitis. *Lancet* 1999;354(9186):1261–1262.
37. Campbell GL, Marfin AA, Lanciotti RS, Gubler DJ. West Nile virus. *Lancet Infect Dis* 2002; 2(9):519–529.
38. Pealer LN, Marfin AA, Petersen LR, et al. Transmission of West Nile virus through blood transfusion in the United States in 2002. *N Engl J Med* 2003;349(13):1236–1245.
39. Detection of West Nile virus in blood donations—United States, 2003. *MMWR Morb Mortal Wkly Rep* 2003;52(32):769–772.
40. http://nationalatlas.gov/virusmap.html
41. www.who.int/csr/sars/casedefinition/en
42. www.who.int/csr/sars/country/table2003_09_23/en
43. Severe acute respiratory syndrome (SARS): Status of the outbreak and lessons for the immediate future. Geneva: World Health Organization. Communicable Disease Surveillance and Response, 2003. pp. 1–10.
44. Ksiazek TG, Erdman D, Goldsmith CS, et al. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 2003;348(20):1953–1966.
45. Rota PA, Oberste MS, Monroe SS, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 2003;300(5624):1394–1399.
46. Fouchier RA, Kuiken T, Schutten M, et al. Aetiology: Koch's postulates fulfilled for SARS virus. *Nature* 2003;423(6937):240.
47. Guan Y, Zheng BJ, He YQ, et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in Southern China. *Science* 2003;302(5643):276–278.
48. Update: multistate outbreak of monkeypox—Illinois, Indiana, Kansas, Missouri, Ohio, and Wisconsin, 2003. *Morb Mortal Wkly Rep* 2003;52(27):642–646.
49. Mwanbal PT, Tshioko KF, Moudi A, et al. Human monkeypox in Kasai Oriental, Zaire (1996–1997). *Euro Surveill* 1997;2(5):33–35.
50. Metzker ML, Mindell DP, Liu XM, et al. Molecular evidence of HIV-1 transmission in a criminal case. *Proc Natl Acad Sci U S A* 2002;99(22):14292–14297.
51. Swofford D. PAUP. Sunderland, Massachusetts: Sinauer Associates, 2001.
52. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 2001;17(8):754–755.
53. Cello J, Paul AV, Wimmer E. Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template. *Science* 2002;297(5583):1016–1018.
54. Jackson RJ, Ramsay AJ, Christensen CD, et al. Expression of mouse interleukin-4 by a recombinant ectromelia virus suppresses cytolytic lymphocyte responses and overcomes genetic resistance to mousepox. *J Virol* 2001;75(3):1205–1210.
55. Wang D, Coscoy L, Zylberberg M, et al. Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci* 2002;99(24):15687–15692.
56. Flint SJ, Enquist LW, Racaniello KR, Skalka AM. *Principles of Virology: Molecular Biology, Pathogenesis, and Control of Animal Viruses. 2nd ed.* Washington, D.C.: ASM Press, 2004.

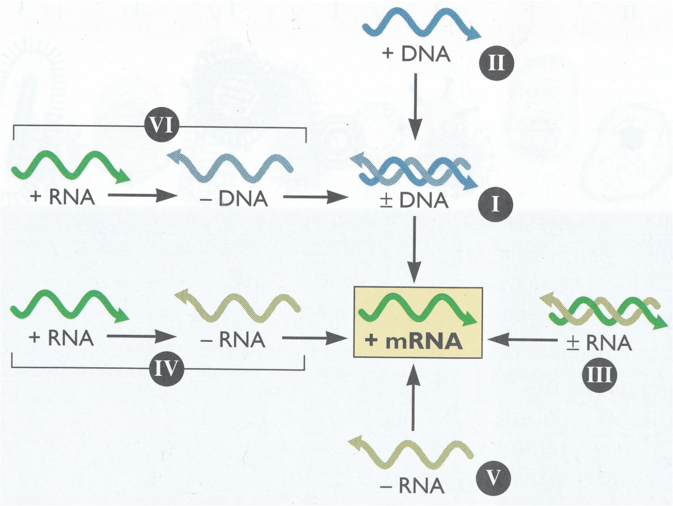

+ DNA
II
VI
+ RNA
− DNA
± DNA
I
+ RNA
− RNA
+ mRNA
± RNA
III
IV
− RNA
V

FIGURE 4.1 Schematic illustrating the replication strategies and means of gene expression for the various Baltimore classes. Reprinted with permission from Ref. (56).

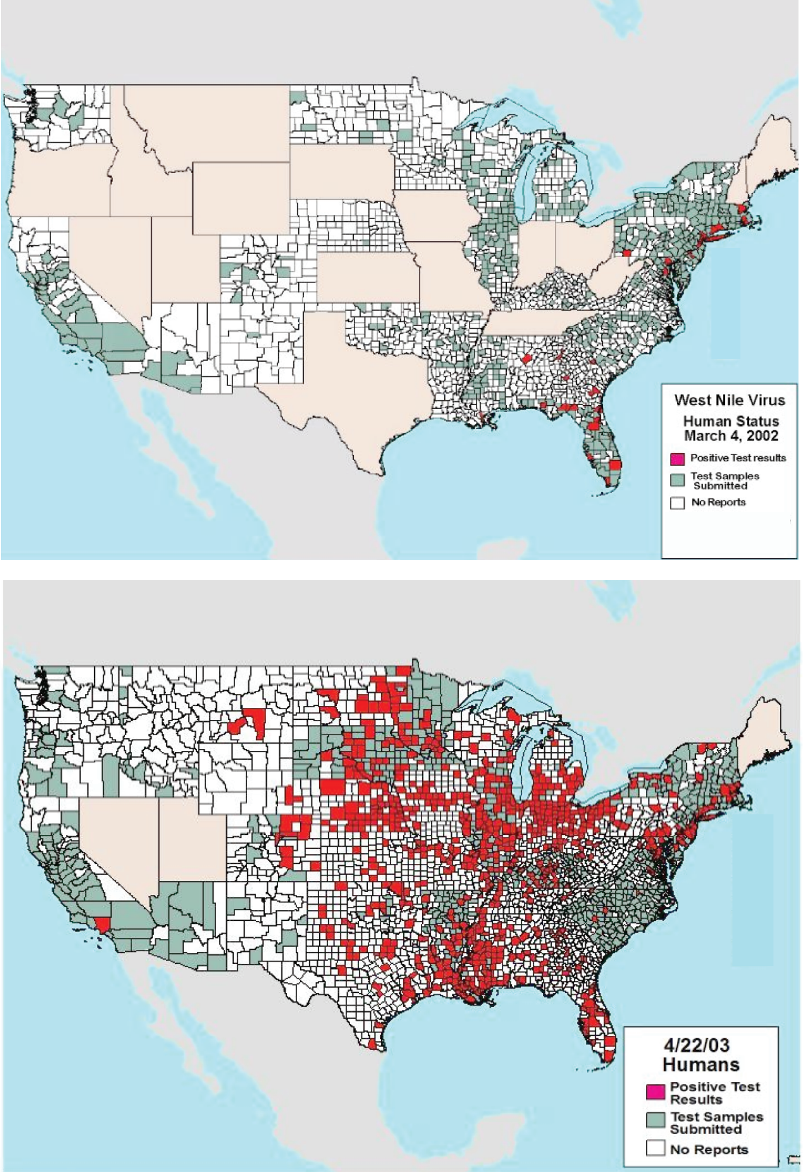

West Nile Virus
Human Status
March 4, 2002
Positive Test results
Test Samples Submitted
No Reports
4/22/03
Humans
Positive Test Results
Test Samples Submitted
No Reports

FIGURE 4.5 Spread of West Nile virus. From the initial outbreak in New York in 1999, the maps chart the rapid westward spread of WNV by showing the geographic distribution of human cases in years 2002–2004. Courtesy of the CDC.

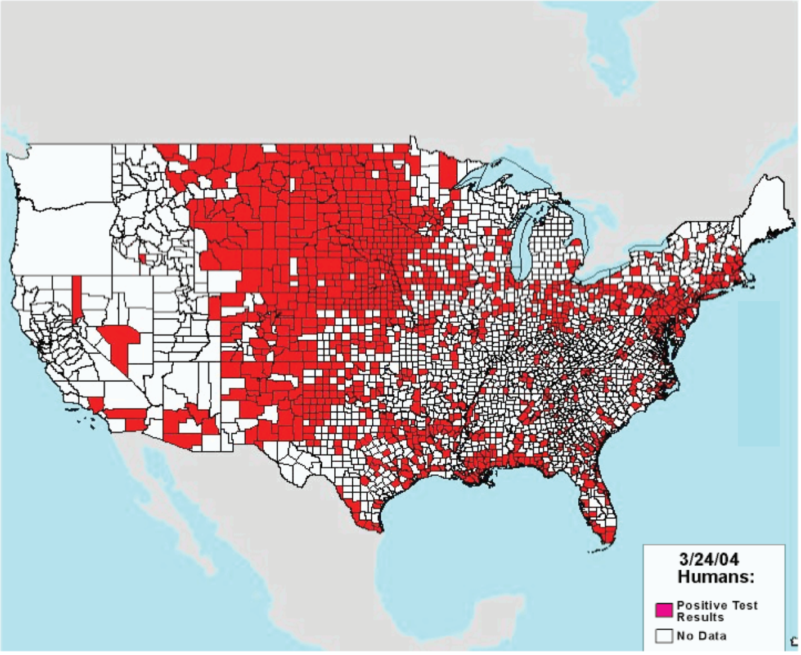

3/24/04
Humans:
Positive Test Results
No Data

FIGURE 4.5 (*Continued*)

CHAPTER 5

Bacterial Pathogens

PAUL KEIM
Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona

Unlike viruses, bacteria are living cells that carry out many complex metabolic functions. Because of this they are larger than viruses, have larger and more complex genomes, and they have cell walls and membranes surrounding a cytoplasm and nucleoid. The cytoplasm is the internal cellular region where many physiological functions occur, and the nucleoid is the cellular region containing a DNA chromosome or chromosomes. Bacterial pathogens have been traditionally divided into two major subdivisions, Gram-positive and Gram-negative, based upon their staining attributes with Gram's stain. This tinctorial differentiation was a useful tool, but provides little taxonomic resolution in this age of sophisticated molecular diagnostics and bacterial systematics. Clearly, forensic scientists need to precisely identify the pathogen to species and if possible identify the particular strain within the species. The more precise the information, the more valuable this will be to investigation, attribution, and the eventual prosecution of the perpetrator. While there are many approaches to such identification, the best current methods rely upon the vast information stored in the bacterial genomes. Extracting this information is readily accomplished by DNA sequencing derived from PCR amplification of particular genomic regions. Different genomic regions have different forensic values in a single bacterial species. These values differ with the pathogen, due to the unique evolutionary and biological properties of each.

DNA REGIONS OF VALUE FOR FORENSIC IDENTIFICATION OF BACTERIAL PATHOGENS

16S rRNA GENE SEQUENCES

Bacterial genomes contain regions that evolve both slowly (e.g., 16S rRNA genes) and very rapidly. Typically, bacterial genomes contain a few million base

Microbial Forensics
Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

pairs of information, most of which is in gene coding regions. There are only small non-gene coding regions separating genes, in stark contrast to eukaryotic genomes. Depending upon the forensic question being asked, different bacterial genomic regions offer unique potential for analysis. If just a general bacterial species or genus identification is needed for the investigation, the 16S ribosomal gene may be the best choice. These are genes coding for an approximately 1500-nucleotide RNA molecule found in the small subunit of the bacterial ribosome. There are great structural constraints upon this molecule which inhibit nucleotide replacement mutations and evolutionary change. Many scientists have contributed to a very large public database of 16S rRNA gene sequences that represents a large fraction of all known bacterial species. Using this database, one can make a very good start towards understanding the diversity within and outside any pathogen target group. In other words, the likely specificity of the 16S rRNA gene sequence should be readily estimable from this database alone. Investigators should be able to determine quickly whether the 16S rRNA gene sequence identifies the target pathogen as distinct and different from other closely related bacteria, while having little variation among different strains of the target. This would be ideal for pathogen target identification and discrimination from non-target bacteria. However, because of its great constancy and slow evolution the 16S rRNA gene sequence frequently does not distinguish the target pathogen from its closest relatives (e.g. *Bacillus anthracis* and some strains of *B. cereus*).

MULTILOCUS SEQUENCE TYPING (MLST)

Housekeeping genes are essential, are found by definition in all bacteria, and evolve more rapidly than 16S rRNA genes. They are thus more discriminating regions. While there are hundreds or even a thousand of these genes available in any genome, researchers tend to pick a small number (less than 10) that are well distributed around the genome. Housekeeping gene regions of value can be studied across the target organism and its close relatives by designing "universal" PCR primers to relatively constant nucleotide sequences to amplify more variable intervening regions. Some pathogens have been extensively studied in this fashion, but many others have not. Hence, forensic investigators will need to generate these data, and the value, or lack of value, in this MLST approach will remain unknown until the studies are completed. However, if an extensive MLST database exists for the pathogen under forensic investigation, this can be a very powerful means to identify the pathogen and subtype strains within the pathogen.

RECENTLY EMERGED PATHOGENS

Some bacterial pathogens have very little genomic variation valuable for subtyping and strain identification. The most commonly invoked cause for this lack of variation is that these organisms have only recently "emerged" as pathogens from a closely related bacterium in the "recent" past. The close bacterial relative may or may not have been a pathogen, but some genetic change presumably occurred (genes could have been gained or lost) that changed its virulence properties and converted it into a highly effective human pathogen. A related but slightly different explanation for genetic uniformity involves a genetic bottleneck whereby an older pathogen population is "stripped" of genomic variation by eliminating all but a single representative of the pathogen. In this scenario, a single pathogenic strain is highly effective and generates all currently known cases of the disease. Whatever the explanation, where there are few or no differences among pathogen isolates, subtyping and precise strain identification are very difficult. This situation is fairly common for bacterial pathogens encountered in biocrimes, and exists in all pathogens that rapidly expand during an epidemic. In other words, lack of genomic variation will be a constant challenge for forensic investigators, who would want the maximum amount of subtyping resolution.

MULTIPLE-LOCUS VNTR ANALYSIS (MLVA)

When the highest subtyping resolution is required for the most precise identification of bacterial isolates (e.g., among lab transfers or cultures), the most diverse and, hence, the most mutable genomic loci must be used. Valuable examples of highly mutable genomic regions are found in the Variable Number Tandem Repeat (VNTR) loci.

VNTR loci are locations where a short nucleotide sequence (e.g., CAT) is repeated multiple times (e.g., . . . CAT CAT CAT CAT . . .). Mutations occur when the cell is replicating its DNA and miscopies the repeats, either increasing or decreasing the copy number. This type of replication error is much more common (~10^5–10^6 higher) than simply miscopying a single nucleotide. The change in repeat copy number is easily detected by designing PCR primers in non-repeated sequences flanking the VNTR and then determining the size of the PCR amplicon by high-resolution electrophoresis. When two pathogen isolates carry identical repeat copy numbers at numerous VNTR loci, it is a very good indication that they are closely related. Likewise, if two isolates are dissimilar at many VNTR loci, it is likely that they are not closely related. Validation of this conclusion must be based upon *in vitro* mutation studies that establish rates for each locus and on population studies that determine the diversity, or discrimination power, at each locus. VNTR loci have been shown to provide strain discrimination power even among closely related isolates of *B. anthracis* and *Y. pestis*.[1,2]

While the genomic presence of VNTR loci varies dramatically among bacterial pathogens, all bacterial genomes appear to have at least a few. VNTR loci's power for discrimination is dictated by individual loci mutation rates, and also by the combined mutation rate across multiple loci (multiple-locus VNTR analysis: MLVA). Monitoring of multiple loci also somewhat mitigates the confounding effect of convergent evolution or mutational reversals at hyper-mutable loci. The forensic utility of MLVA depends upon:

1. Genomic identification of particular VNTR loci for each pathogen,
2. Development of efficient assays (e.g. PCR),
3. Population studies that describe the pathogen diversity for these assays, and
4. Mutational studies that lead to mathematical evolutionary models.

Population and mutational studies must be repeated for each pathogen and its VNTR loci. While some generalization is feasible among species, bacteria are highly diverse organisms and each should be considered a unique entity. In extremely critical situations, mutational studies should be carried out with the isolates actually involved in the forensic scenario under investigation to eliminate any possibility of strain-specific effects.

BACTERIAL PATHOGENS OF CONCERN

Many different bacterial pathogens can and have been employed in biocrimes. It is difficult to predict which pathogens will be used in the future, although availability, ease of handling, ease of weaponization, and potential impact (death or terror) are factors that might narrow the options.[3] Traditionally bioterrorism defense priorities and pathogen-target lists have been greatly influenced by bacterial agents that were part of "state-sponsored" bioweapons programs. Thus, *Bacillus anthracis* (anthrax), *Yersinia pestis* (plague), *Francisella tularensis* (tularemia), *Burkholderia* sp. (glanders and melioidosis), and *Brucella* sp. (brucellosis) are often cited as possible bioterrorism agents. The 2001 anthrax letter attacks demonstrated that this is a very relevant list of potential biocrime agents. However, other commonly available pathogens must also be considered. In the U.S., two notable biocrimes used the food-borne pathogens *Shigella* and *Salmonella*. The very common occurrence of these diseases means that the pathogenic agents are readily available. In addition, the

U.S. food supply system is readily accessible to criminals, making such pathogens important considerations for microbial forensics.

Numerous websites provide detailed information on these priority pathogens, the diseases they cause, and the current status of genomic analysis. The following examples provide the reader with a starting point:

- http://www.hc-sc.gc.ca/pphb-dgspsp/, Health Canada
- http://www.cdc.gov/ncidod/dbmd/diseaseinfo/, Centers for Disease Control and Prevention
- http://www.idph.state.ia.us/eedo/cade_content/epifacts/, Iowa Department of Public Health
- http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html, National Center for Biologic Information
- http://www.tigr.org/tdb/mdb/mdbinprogress.html, The Institute for Genomic Research

BACILLUS ANTHRACIS

The causative agent of anthrax is a low-GC, Gram-positive, spore-forming bacterium. The spore is a definitive feature of its ecology, evolution, and use as a bioweapon. Particular strains of *B. cereus* and *B. thuringensis* isolates have been shown to be the closest relatives to *B. anthracis*. Because these related species are common in the environment and as industrial microbes, differentiation between them and *B. anthracis* is crucial. Based upon 16S rRNA sequences, there are few, or no, differences between *B. anthracis* and its closest relatives. However, *B. anthracis* contains two large plasmids that carry essential virulence genes for protein toxins (pXO1) and the biosynthesis of a poly-D glutamic acid capsule (pXO2). Lack of either plasmid dramatically attenuates the bacterium.

B. anthracis is a zoonotic disease and probably pathogenic to many mammals. However, its virulence differs greatly among the species—herbivores are most commonly affected in nature. The frequency of the disease in herbivores could be the result of ingestion of spores along with their food, followed by invasion from the gastrointestinal tract. Humans most commonly exhibit cutaneous anthrax, acquired by handling spore-contaminated wool or animal hides. Gastrointestinal anthrax from eating spore-contaminated meat is the next most common form, with inhalational anthrax being relatively rare. There is no evidence of strain differences between infection routes.

Because of the historical and recent importance of *B. anthracis* as a biological weapon, genomic and genetic analyses are highly developed for this pathogen. Multiple strain genomes have been completely sequenced, including the Ames isolate from the first anthrax-letter victim of the 2001 U.S. attacks.[4,5] High-resolution DNA genotyping based upon MLVA can uniquely identify most laboratory strains, with some differentiation possible among different isolates of the same strain.[1]

YERSINIA PESTIS

This Gram-negative enteric bacterium causes plague. Its closest relatives are the more mildly pathogenic *Yersinia pseudotuberculosis* and *Y. enteroclitica* (other well known enteric bacterial pathogens include *Escherichia coli* and Salmonella species). Three biovars of *Y. pestis* are diagnostically differentiated by the ability of an isolate to ferment glycerol and reduce nitrate. The three biovars (antigua, medieavalis, and orientalis) are now endemic in particular regional geographic foci thought to be the remnants of three major plague pandemics.

Y. pestis is a zoonotic pathogen thought to reside in animal reservoirs. The two recognized clinical forms of human plague—pneumonic and bubonic—are different pathological manifestations of infection with the same pathogen. An important distinction is the transmission route, since pneumonic plague can be extremely contagious. Bubonic plague is more commonly associated with bites of insect vectors, especially fleas. In different global locations, endemic plague reservoirs may be maintained by low-susceptible animal hosts, quiescent insect vectors, or a slower transmission cycle in otherwise susceptible host populations. Worldwide, endemic plague foci are locally established, and the reservoir ecology is probably unique in each case. In addition to humans, many mammalian species are known to be variably susceptible to plague.

Plague genomics and genotypic analysis are advanced. Two high-quality finished genomes are publicly available at this time (Kim and CO92),[11] and hypermutable loci have been identified for high-resolution genotyping.[2,6]

FRANCISCELLA TULARENSIS

This Gram-negative coccobacillus causes tularemia in both the Old and New Worlds. Two subspecies (biovars) have different global distributions. *F. tularensis* ssp. *tularensis* (biovar A) and *F. tularensis* ssp. *holartica* (biovar B) are both found in North America; biovar B is found in Eurasia; biovar A is very rare in Eurasia. The severity of tularemia due to biovar A is much greater than that of biovar B. Though the genetic basis for this difference is not currently understood, it is under active research. *F. tularensis* survives well in the environment and persists on contaminated meat, water, hides, and in vectors for weeks or even years. Tularemia can be acquired from arthropod vectors such as deer flies and ticks. *F. tularensis tularensis* in North American rabbits is extremely virulent. In contrast, *F. tularensis holartica* infects water and aquatic mammals (e.g., beavers) with much less virulence.

High-resolution DNA subtyping has been described,[7,8] and two studies are underway to define the genome sequences of each biovar (LVS and Shu4).[12] Comparison of the genomes of the two subspecies should provide insights into their unique virulence properties. MLVA analysis of diverse strains was consistent with the traditional biovar categories and uniquely identified most isolates, but was unable to discriminate among epidemiologically linked isolates.[7] The genetic identity of isolates from clustered disease cases is consistent with these being initiated via a single infectious event. MLVA analysis also indicated that *F. tularensis tularensis* was much more diverse, suggesting an older age for this subspecies, or that the studied *F. tularensis holartica* isolates had been through a great population bottleneck.

BRUCELLA

Multiple species of Brucella, all Gram-negative bacteria, have been differentiated primarily through their association with particular hosts. The three public health and forensic microbiology priorities are *B. melitensis*, *B. suis*, and *B. abortus* (in order of importance). *B. melitensis* is the most common and pathogenic. Brucellosis is most commonly transmitted through contaminated milk or cheese that has not been pasteurized, although direct contact with infected animals can cause occupational disease. Brucellosis is a serious health risk in less developed countries where hygienic standards are low and milk products are not pasteurized. In developed countries, Brucellosis is the most commonly reported laboratory-acquired infection due, perhaps, to its environmental stability and aerosol infectivity.

Two *Brucella* genomes have been completed and others are in process. Both the *B. melitensis*[9] and the *B. suis*[10] genomes are separated into two chromosomes of about 2 million and 1 million nucleotides. High-resolution VNTR analysis systems under development appear capable of precise strain identification (L. Huynh and P. Keim, unpublished data).

MELIOIDOSIS AND GLANDERS

The pathogens responsible for melioidosis (*Burkholderia pseudomallei*) and glanders (*B. mallei*) are two closely related species. In fact, recent molecular analysis appears to indicate that *B. mallei* is a clonal derivative of the much more diverse *B. pseudomallei*. Both are Gram-negative aerobic rods, but cause very different diseases. Glanders is now an uncommon disease of horses and donkeys. Its natural occurrence is restricted to the Indian subcontinent, Africa, some Middle Eastern countries, and parts of eastern Asia. In North America and Europe, human glanders cases are primarily associated with high-risk laboratory or animal facility workers. The only contemporary human case in the U.S. was the result of an accidental laboratory exposure. *B. mallei* does not survive outside susceptible host animals. In contrast, melioidosis occurs worldwide, though primarily in tropical and subtropical regions. Melioidosis is uncommon in the U.S. except in returning travelers or immigrants. However, it is endemic in Southeast Asia and elsewhere. *B. pseudomallei* survives well in the environment (moist soil and water) to provide ample sources for human and animal infection.

Whole genomes of both *B. mallei* and *B. pseudomallei* are being characterized. They have a bipartite genomic structure with both a primary and secondary chromosome. The great diversity of *B. pseudomallei* suggests that multiple strains will need to be analyzed to understand this pathogen. Both gene nucleotide sequencing and hypervariable loci have demonstrated the high diversity found among isolates of *B. pseudomallei*. This may be due to a highly plastic genome and the ecological and environmental niche of the bacterium. Development of high-resolution DNA typing is underway, and precise identification will be possible due to the great diversity of strains (J. Schupp and P. Keim, unpublished data).

REFERENCES

1. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ, Hugh-Jones ME. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol* 2000;182:2928–2936.
2. Klevytska AM, Price LB, Schupp JM, Worsham PL, Wong J, Keim P. Variable number tandem repeat diversity is great and genomically widespread in *Yersinia pestis*. *J Clin Microbiol* 2000;38:3179–3185.
3. Keim P. Microbial forensics: A scientific assessment. Washington, DC: Am Acad Microbiol; Feb. 17, 2003; Report No. DOEER63403.
4. Read TD, Peterson SN, Tourasse N, et al. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* 2003;423:81–86.
5. Read TD, Salzberg SL, Pop M, Shumway M, Umayam L, Jiang L, Holtzapple E, Busch JD, Smith KL, Schupp JM, Solomon D, Keim P, Fraser CM. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* 2002;296:2028–2033.
6. Adair DM, Worsham PL, Hill KK, Klevytska AM, Jackson PJ, Friedlander AM, Keim P. Diversity in a variable-number tandem repeat from *Yersinia pestis*. *J Clin Microbiol* 2000; 38:1516–1519.

7. Farlow J, Smith KL, Wong J, Abrahms M, Keim P. *Francisella tularensis* strain typing using variable number tandem repeats. *J Clin Microbiol* 2001;39:3186–3192.
8. Johansson A, Goransson I, Larsson P, Sjostedt A. Extensive allelic variation among *Francisella tularensis* strains in a short-sequence tandem repeat region. *J Clin Microbiol* 2001;39: 3140–3146.
9. DelVecchio VG, Kapatral V, Redkar RJ, et al. The genome sequence of the facultative intracellular pathogen *Brucella melitensis*. *PNAS* 2002;99:443–448.
10. Paulsen IT, Seshadri R, Nelson KE, et al. The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. *PNAS* 2002;99:13148–13153.
11. Deng W, Burland V, Plunkett G III, et al. Genome sequence of *Yersenia pestis* KIM. *J Bacteriol* 2002;184:4601–4611.
12. Parkhill J, Wren BW, Thomson NR, et al. Genome sequence of *Yersenia pestis*, the causatire agent of plague. *Nature* 2001;467:469–470.

CHAPTER 6

Biology and Detection of Fungal Pathogens of Humans and Plants

Gary A. Payne and Marc A. Cubeta
Department of Plant Pathology, North Carolina State University, Raleigh, North Carolina

Fungi represent an important group of pathogens that significantly impact human and plant health. They are responsible for the majority of plant diseases and are of increased importance as agents of infectious diseases of immunocompromised humans. Most human and plant pathogens occur in the phylum Ascomycota and are represented in taxonomically distinct classes within the phylum. There are relatively few examples of fungi that are both human and plant pathogens. Disease is the result of an interaction between a pathogen and a susceptible host within a favorable environment. The interrelationships associated with disease are complex and not well characterized. Factors required for pathogenesis include toxins, enzymes, and signaling pathways associated with both fungal morphology and the molecular communication between the host and pathogen. Fungi produce abundant numbers of spores that are well adapted for short- and long-distance dispersal from infected to healthy plants. Repeated dispersal and cycles of infection on the same plant are common attributes of many plant pathogens. In contrast, fungal diseases of humans are usually not communicable, and not transmitted from an infected to a healthy individual. Polymerase chain reaction (PCR)- and DNA sequence-based methods are allowing for a closer examination of the ecology and epidemiology of human and plant pathogens. These methods are increasing our understanding of the spread of infectious diseases, and provide pragmatic approaches for rapid detection and identification of indigenous and introduced fungal pathogens. This chapter provides a review of fungi, including their taxonomy, growth, reproduction, pathogenicity, epidemiology, detection, and identification.

Microbial Forensics
Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

INTRODUCTION

Fungi represent a diverse assemblage of eukaryotic organisms and are common inhabitants of most aquatic and terrestrial ecosystems, where they are of significant ecological and economic importance. Approximately 100,000 species of fungi have been described, and it is estimated that 1.5 million are in existence.[1] Along with bacteria, fungi are largely responsible for the decomposition and recycling of organic matter, particularly recalcitrant molecules of cellulose and lignin. Fungi play a pivotal role in determining the composition and diversity of terrestrial ecosystems, because the roots of most land plants (>90%) have an established mycorrhizal association with at least one species of fungi.[2] In addition, approximately 20% of all fungi are involved in a symbiotic partnership with green algae or cyanobacteria (previously referred to as blue-green bacteria) to form lichens.[3]

Fungi have greatly shaped the history of humankind. For example, the yeast *Saccharromyces cerevisiae* has been cultured for centuries for the production of breads and alcohols. Likewise, several species of *Aspergillus* have been used for thousands of years in Asia for the production of miso, tempeh, and soy sauce. Fungi can also be eaten directly as a food source (e.g., blue cheese, morels, mushrooms, truffles, and the recently marketed product *Quorn*®, a mycoprotein from the filamentous fungus *Fusarium venenatum*).[4] These culinary and nutritional uses along with more recent employment of fungi for the production of industrially important enzymes (laccases, cellulases, proteases, xylanases), organic acids (citric, gluconic, itaconic), antibiotics (cephalosporin, penicillin), and pharmaceuticals (cyclosporin, lovastatin) highlight our dependence on fungi.[5–11]

Not all fungi are beneficial. Some fungi are harmful pathogens and can cause diseases of animals, humans, and plants. At least 8,000 species of fungi are known to cause plant diseases, and less than 100 species are known pathogens of humans.[12,13] More information on human and plant pathogenic fungi can be found at the following websites: http://nt.ars-grin.gov/sbmlweb/OnlineResources/FungalDBsDescrip.cfm and http://timm.main.teikyou.ac.jp/pfdb/cover/database_eng.html.

Many fungi contaminate food by producing toxic secondary metabolites known as mycotoxins. Mycotoxins can produce a wide array of symptoms in animals and humans. Every system of the human body can be affected by at least one mycotoxin.[14] More than 300 mycotoxins have been described, and potentially many more exist. It has been estimated that more than 200,000 secondary metabolites are produced by fungi, most of which have not been characterized and assayed for direct toxicity.[14] Some examples of mycotoxins are aflatoxins, ergot alkaloids, fumonisins, trichothecenes, and zearalenone. Although not usually considered a mycotoxin, the amanitin toxins produced by several poisonous species of *Amanita* (e.g., the death cap *A. phalloides* and destroying angel *A. virosa* mushrooms) inhibit the RNA polymerase II enzyme associated with gene transcription and are often lethal to humans if these mushrooms are consumed.[15]

CLASSIFICATION, NOMENCLATURE, AND TAXONOMY OF FUNGI

For many centuries fungi were thought to be closely related to plants and were not classified in their own Kingdom until 1969.[16] Recent analyses of DNA and protein sequences in fungi suggest that the Kingdom Fungi is a sister group to the Kingdom Animalia, providing convincing evidence that fungi are more closely related to animals than plants.[17] Interestingly, whole-genome analysis has revealed significant similarity among some plant and fungal genes.[18] For more detailed information on the phylogenetic relatedness of fungi to other living organisms, see the Tree of Life website http://tolweb.org/tree/phylogeny.html.

Organisms classified in the Kingdom Fungi include four phyla: Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota, and are often referred to as the "true fungi" or Eumycota.[2] Typical features of the true fungi include filamentous growth, cell walls that contain chitin, a predominant haploid life cycle, flattened plate-like mitochondrial cristae, and absorptive nutrition (i.e., the ability to digest, rather than ingest, nutrients for their growth and development). Historically, several evolutionarily divergent groups of organisms such as the slime molds (Mycetozoans) and members of the phylum Oomycota (water molds) have been included in the Kingdom Fungi. However, recent molecular evidence based on analysis of DNA and protein sequence data suggests that these organisms are not phylogenetically related to the true fungi. Organisms classified in the Oomycota are of particular ecological and economic importance because they can cause diseases of amphibians, animals, fish, and plants. Members of the Oomycota have often been described as being "fungus-like" because of they have mycelium and exhibit filamentous growth and absorptive nutrition. However, organisms in the Oomycota are unlike true fungi in that they have cellulose-containing cell walls, a predominant diploid life cycle, different-shaped mitochondrial cristae, and a different lysine biosynthesis pathway.[19] Because of these differences, organisms classified in the phylum Oomycota have been recently placed into the newly proposed Kingdom Straminipila that also includes brown algae and diatoms.[20] Members of the Kingdom Straminipila (also referred to as "straminipiles") are also characterized by the production of swimming spores called zoospores with usually two tail-like appendages called flagella (singular, flagellum) that are used for locomotion. One flagellum is hairless, while the other flagellum is decorated by two rows of tripartite tubular hairs. The latter is often referred to as a heterokont or tinsel flagellum.[20] The tinsel flagellum is the basis for the term straminipile (which literally translates to "straw hairs") and for naming the Kingdom Straminipila.

Fungi were among the first microorganisms observed. In 1664, Robert Hooke was able to visualize the spores of a "blue mold fungus" on a piece of leather with a microscope.[21] Since this initial observation, scientists interested in the scientific study of fungi (i.e., mycologists) have traditionally utilized morphological characteristics (mostly color, shape, and size) of asexual and sexual spores and how they form (ontogeny), for naming (nomenclature) and classifying (taxonomy) fungi. Fungal spores can be colorless (hyaline) or various shades of the visible spectrum of light, including brown and black. Spores that are darkly pigmented (olive-, brown-, or black-colored) because of the accumulation of melanin in their cells are referred to as being dematiaceous.

In general, most species of fungi produce at least one type of spore during their life. However, some fungi do not produce any spores, while others have the capability of producing more than one form or type of spore during their life (a phenomenon referred to as dimorphism or pleomorphism). These complex characteristics of fungal life history have provided challenges for fungal nomenclature and taxonomy. In many instances, the same fungus can have two different scientific names that correspond to its asexual and sexual stages. For example, the common soil-inhabiting fungus *Rhizoctonia solani* is the asexual stage of *Thanatephorus cucumeris*. This method of dual taxonomy represents a unique aspect of nomenclature in the biological world. Although the taxonomy and nomenclature of fungi have always been in a constant state of flux, it serves as an effective mechanism to foster communication among scientists to increase our knowledge and understanding of fungi.

A variety of biochemical (fatty acids, enzymes, and proteins), ecological (habitat and pathology), genetic (vegetative compatibility and mating types), and physiological (secondary metabolites) characteristics are commonly employed for fungal nomenclature and taxonomy. However, morphological characteristics of the sexual stage still represent the primary basis for the classification of fungi, although many fungi are asexual and do not have a sexual stage. Despite these shortcomings, fungi are separated into groups with similar characteristics based on the following principal hierarchy: (1) species grouped into a genus, (2) genera grouped into a family, (3) families grouped into an order, (4) orders grouped into a class, (5) classes grouped into a phylum, (6) phyla grouped into a Kingdom, and (7) Kingdoms grouped into a biological Domain (Figure 6.1).

Taxonomic rank	Suffix
Domain	Eukarya
Kingdom	Fungi
Phylum	-mycota
Class	-mycetes
Order	-ales
Family	-aceae
Genus	
Species	

FIGURE 6.1 Principal taxonomic ranks and suffixes used in fungal nomenclature.

Recent phylogenetic analyses based on analysis of ribosomal DNA (rDNA) and protein gene sequence data support the existence of four monophyletic groups within the Kingdom Fungi. A brief description of the major phyla in the Kingdom Fungi and their general characteristics is presented below.

CHYTRIDIOMYCOTA (CHYTRIDS)

Organisms in this phylum represent the oldest branch (most basal clade) of true fungi; there is fossil evidence of their existence approximately 400 million years ago.[22] Chytrids produce motile zoospores with usually a single "hairless" flagellum and lack filamentous mycelium. Their cell walls are similar to those of other fungi, insects, and arthropods and are composed primarily of polymers of chitin (N-acetyl glucosamine) and glucans (polysaccharides composed of solely or partly glucose molecules joined through α-1,3; α-1,6; β-1,3; or β-1,6 linkages). β-glucan mixtures of β-1,3 and β-1,6 linkages represent the most abundant cell wall component in fungi, and ergosterol is also present in cell membranes of fungi.

Chytrids are present in most habitats where water is present, including in soil, and can be damaging pathogens of algae, amphibians, fungi, plants, and small animals.[23] However, only a few chytrids are pathogens of economically important vascular species of plants. For example, *Synchytrium endobioticum* causes the devastating wart disease of potato and was classified as a quarantine pathogen in the Plant Protection Act enacted by the U.S. Department of Agriculture in 1912. Organisms in the genus *Neocallimastix* represent a unique group of chytrids because they grow in the absence of oxygen (i.e., they are anaerobic) within the rumen of cows and sheep to facilitate the degradation of cellulose. This mutualistic association between the animal and fungus provides much needed energy and nutrients for their growth and development.

ZYGOMYCOTA (BREAD MOLDS)

Organisms in this phylum are largely responsible for food spoilage and are usually weak pathogens of plants. They exist primarily as saprophytes in dung, plant debris, and soil. However, some members of the Zygomycota are important pathogens of amoebae, humans, insects, and nematodes. The human pathogens are usually associated with diseases of immunocompromised individuals (zygomycosis), while other species can also cause tropical subcutaneous mycoses in noncompromised individuals. Many species in the Zygomycota (e.g., *Gigospora*, *Glomus*, and *Scutellospora*, etc.) are involved in mutualistic associations within the roots (i.e., endomycorrhizal) of more than 70% of the species of herbaceous plants.[2] Depending on structures formed within the roots of plants, endomycorrhizae are classified as being vesicular arbuscular mycorrhizal (VAM), or arbuscular mycorrhizal (AM) fungi.

Members of the Zygomycota exhibit filamentous growth, and hyphal cells that comprise the mycelium are not delimited by cross-walls (e.g., aseptate, coenocytic mycelium). The hyphal cell walls are composed predominantly of chitosan and polyglucuronic acid, but chitin is also present. Unlike the chytrids, members of the Zygomycota produce nonmotile, asexual (sporangiospore), and sexual (zygospore) spores. Recent evidence suggests that the Zygomycota are evolutionarily divergent, and that this phylum may not represent a monophyletic group of fungi derived from a single common ancestor.[22,24]

ASCOMYCOTA (SAC FUNGI)

Organisms in this phylum include important pathogens of animals, humans, insects, and plants, usually with a saprophytic stage during their life cycle. Mutualistic associations with green algae and cyanobacteria (lichens) and vascular plants (ectomycorrhizae) are also known and of considerable ecological importance. It has been estimated that more than 40% of lichen forming fungi belong to the phylum Ascomycota.[3] This phylum also includes the morels, truffles, and yeasts. Members of the Ascomycota can exhibit single-celled (yeasts) and/or filamentous growth. The hyphal cells that form the mycelium are compartmentalized by cross-walls (e.g., septate, noncoenocytic mycelium), and their cell walls are composed of chitin and β-glucans. Similar to the

Zygomycota, members of the Ascomycota usually produce nonmotile asexual spores called conidia (singular, conidium) and/or sexual spores called ascospores. The ascospores are produced in a sac-like structure called an ascus (plural, asci) and usually dispersed by wind.

BASIDIOMYCOTA (CLUB AND MUSHROOM FUNGI, RUSTS, AND SMUTS)

Organisms in this phylum share many characteristics with the Ascomycota, and recent analyses of molecular and morphological data suggest that the Ascomycota and Basidiomycota are sister groups in the Kingdom Fungi.[25] Some common characteristics shared by members of the Ascomycota and Basidiomycota include (1) pathogens of animals, humans, insects, and plants; (2) mutualistic associations with roots of woody species of plants (ectomycorrhizal); (3) single-celled, filamentous, and saprophytic growth; (4) septate, noncoenocytic mycelium with cell walls composed of chitin and β-glucans; and (5) production of nonmotile asexual and sexual spores. In contrast to fungi in the Ascomycota which produce their sexual spores in an ascus, fungi in the Basidiomycota produce their sexual spores (basidiospores) on a club-shaped structure called a basidium (plural, basidia).

Because many fungi do not produce sexual spores, another phylum called the Deuteromycota has been designated to classify fungi. More than 1,700 genera and 17,000 species of fungi belong to the Deuteromycota and likely represent the asexual stages of fungi in the Ascomycota and Basidiomycota.[2] With the advent of molecular techniques it is now possible to determine the phylogenetic relatedness of asexual fungi to their sexual relatives. At the present time, the phylum Deuteromycota provides a pragmatic approach for identifying fungi that do not undergo sexual reproduction.

GROWTH HABIT AND REPRODUCTION

The life cycle of a fungus can vary considerably among species with respect to growth habit and reproduction. In general, most fungi produce some type of asexual and/or sexual spore during their life cycle that enables them to grow, be disseminated, and/or survive. Under favorable environmental conditions, spores germinate to produce a thread-like filament called a primary hypha (plural, hyphae) or germ tube. Hyphae grow by extension at the hyphal tips in a branching pattern that allows maximum absorption of nutrients from a substrate. The hyphae become organized into the vegetative body of the fungus called a mycelium (plural, mycelia). Fungi are nonphotosynthetic and rely on

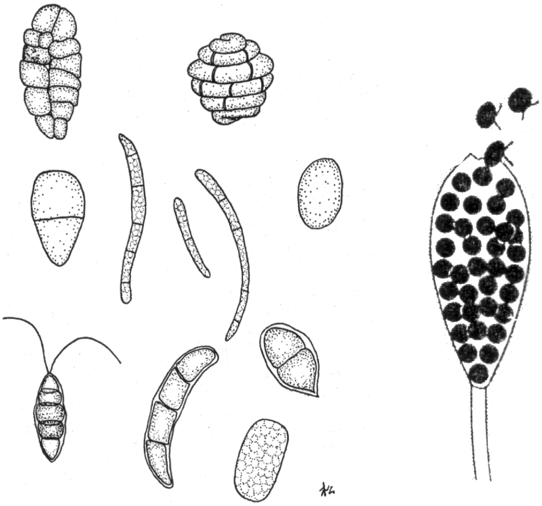


FIGURE 6.2 Asexual spores produced by fungi. Structure to the far right is a sporangium that contains and is releasing motile zoospores (not to scale). Modification of Fig. 2.25 in ref. (2). Printed with permission from John Wiley and Sons.

absorption to obtain their nutrition. They produce an array of extracellular degradative enzymes necessary to break down complex molecules from dead and/or living cells into simpler compounds that can be readily absorbed.

As fungi colonize their substrate, spores may be produced via asexual or sexual reproduction. Spores produced via asexual reproduction develop as a result of mitosis without the fusion of gametes or nuclei. In contrast, spores produced via sexual reproduction involve the fusion of nuclei and/or gametes via meiosis. Examples of asexual spores include aleuriospore, chlamydospores, conidia, sporangiospores, urediospores, and zoospores (Figure 6.2). Asexual spores can form singly on a hyphal cell or form in groups into a specialized structure (e.g., synnema, sporodochium, Figure 6.3). Asexual spores can also be produced on hyphal cells in an enclosed or partially enclosed fruiting structure (pycnidium, acervulus, Figure 6.3).

The morphology and ontogeny of asexual spores (conidia) provide a basis for their classification in the Deuteromycota. Multiple generations of asexual spores may be produced in a given year depending on environmental conditions and host susceptibility. Asexual spores are important for rapid dissemination and propagation of identical genetic individuals of a pathogen and for subsequent epidemic development. Although sexual spores can have functions similar to those of asexual spores, they are also important for long-term survival and for serving as a mechanism for generating genetic diversity

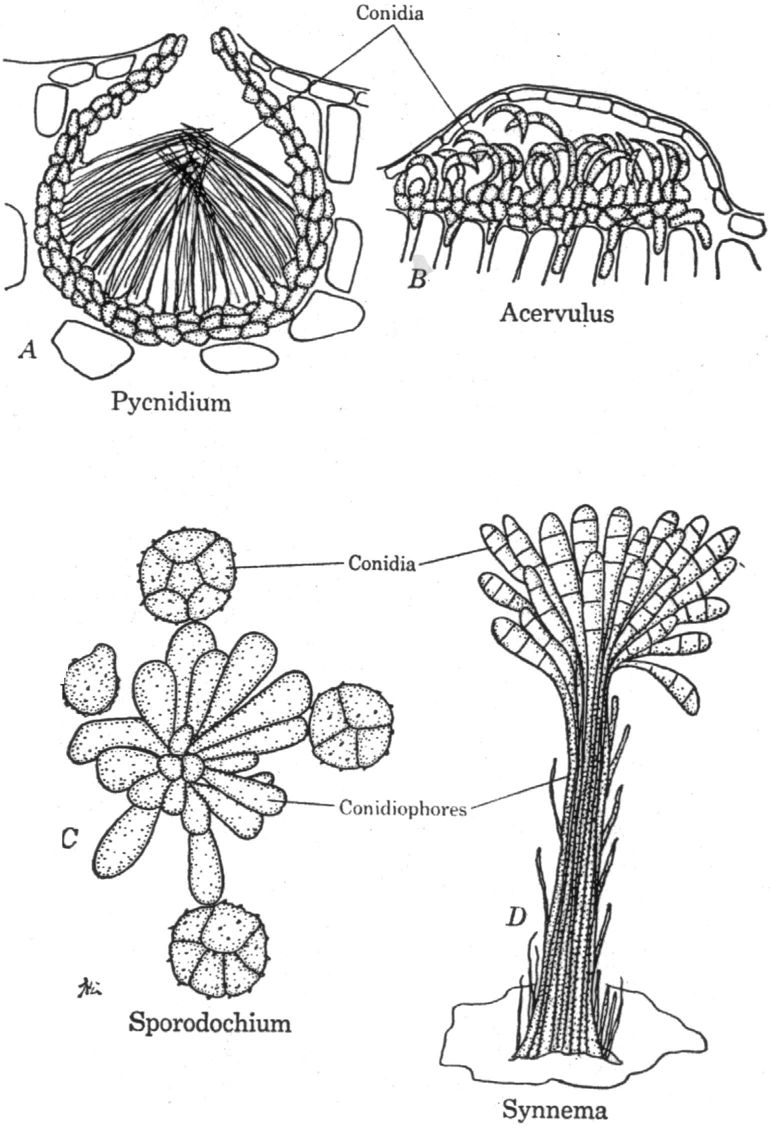

Conidia
A
Pycnidium
B
Acervulus
Conidia
Conidiophores
C
Sporodochium
D
Synnema

FIGURE 6.3 Fruiting structures associated with asexual fungal spores (conidia). Fig. 8.3 from ref. (2). Printed with permission from John Wiley and Sons.

through recombination. Fungi can also produce specialized structures such as microsclerotia or sclerotia (a compacted mass of melanized cells) to enhance their survival during periods of unfavorable temperature and moisture conditions. Examples of sexual spores include aeciospores, ascospores, basidiospores, oospores, sporidia, and zygospores. In general, ascospores are produced in a specialized fruiting structure called an ascocarp. Four different types of ascocarps have been described: apothecium, cleistothecium, perithecium, and pseudothecium (Figure 6.4). However, some fungi in the phylum Ascomycota do not produce ascospores in an ascocarp. Basidiospores are produced on a basidium that is or is not enclosed in a fruiting body or mushroom (basidiocorp), while oospores and zygospores are usually borne directly on or in hyphal cell (Figures 6.5 and 6.6).

For more detailed information see the websites by Tom Volk, http://botit.botany.wisc.edu/toms_fungi/, and Kathy Hodge, http://biodiversity.uno.edu/~fungi/.

PATHOGENICITY

Growth and reproduction are fundamental aspects of the fungal life cycle, and they are important for pathogenic invasion within animals and plants. In general, the majority of species of fungi are saprophytes or parasites, but not pathogens. A parasite is an organism that acquires nutrition from a living host, but the term parasite does not imply a beneficial or harmful interaction. In contrast, a pathogen acquires nutrients at the expense of the host, and this interaction results from physiological changes in the host and subsequent appearance of disease symptoms. At least 8,000 and 100 species of fungi are known pathogens of plants and humans, respectively.[13,26]

Each of the four phyla in the Kingdom Fungi has pathogenic species, but the Ascomycota contains the largest number of pathogens. Within the Ascomycota, most pathogens of plants and animals are restricted to a few taxonomic classes.[27] In general, plant pathogens are represented in different classes than animal pathogens. A large percentage of plant pathogens are members of the classes Taphrinomycetes, Sordariomycetes, Leotiomycetes, and Dothideomycetes, whereas animal pathogens belong mostly to the Eurotiomycetes and Chaetothyriomycetes.[27] In addition, the Basidiomycota, Chytridiomycota, and Zygomycota also contain important pathogens of animals and plants.[28,29]

Most plant pathogenic fungi can infect and colonize healthy hosts. In contrast, the majority of fungi that cause diseases of humans infect hosts who are debilitated or compromised, and these fungi are known as opportunistic pathogens. However, two exceptions are *Coccidioides immitis* and *Histoplasma capsulatum*,[30] which infect apparently healthy individuals.

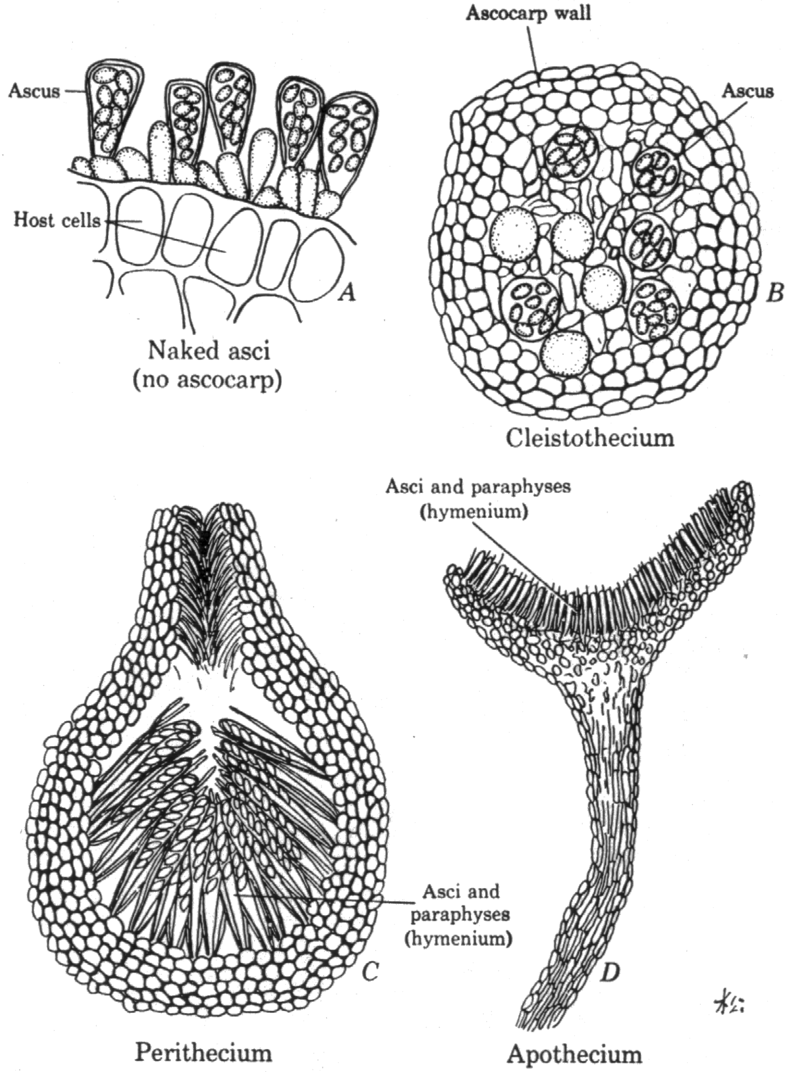

Ascocarp wall
Ascus
Host cells
A
Naked asci
(no ascocarp)
Ascus
B
Cleistothecium
Asci and paraphyses
(hymenium)
Asci and
paraphyses
(hymenium)
C
D
Perithecium
Apothecium

FIGURE 6.4 Fruiting structures (ascocarps) produced by fungi in the phylum Ascomycota with asci and ascospores; (A) Naked asci, (B) Cleistothecium, (C) Perithecium, and (D) Apothecium. Fig. 11.11 from ref. (2). Printed with permission from John Wiley and Sons.

Over the last 20 years the steady increase in the number of human infections caused by opportunistic fungi has presented a greater concern to the medical community. This increase in human disease is associated with opportunistic fungal infections in compromised individuals who have AIDS or have

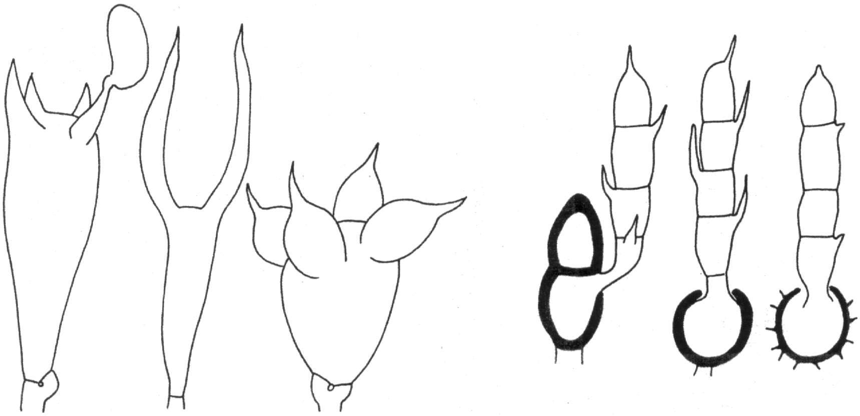


FIGURE 6.5 Different types of basidia produced by fungi in the phylum Basidiomycota. Note basidiospore on tip of sterigma on basidium on the far left. The three basidia to the far right are typical of those produced from teliospores of rust and smut fungi. Modification of Fig. 11 ref. (1). Printed with permission from CAB International.

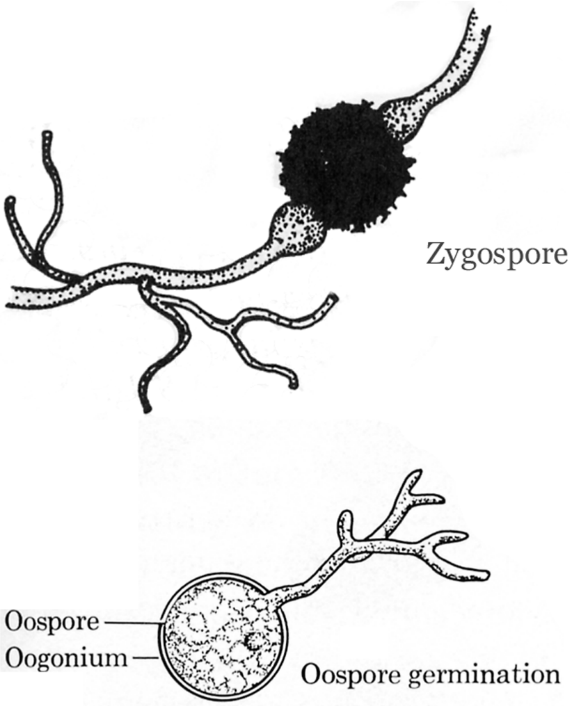


FIGURE 6.6 Zygospore and oospore produced by organisms in phyla Zygomycota and Oomycota, respectively (not drawn to scale). Modification of Fig. 10 from ref. (19), and Fig. 8.18 from ref. (2). Printed with permission from John Wiley and Sons.

undergone medical procedures that require immunosuppression, e.g., bone marrow and organ transplants and chemotherapy. Examples of opportunistic pathogens are *Cryptococcus neoformans*, *Aspergillus fumigatus*, species of *Candida*, *Aspergillus*, *Fusarium*, *Penicillium*, and several species within the Zygomycota. In the last few years the number of invasive infections by species of *Aspergillus* has risen steadily, replacing species of *Candida* as the mostly commonly encountered fungal pathogen in some institutions.[31] Further, species of *Fusarium* are now recognized as second to species of *Aspergillus* as a major invasive mold.[31] Also, species of *Candida* other than *C. albicans* are becoming increasingly more prevalent, possibly due to the extensive use of antifungal drugs toxic to *C. albicans*.[31]

The interrelationship between a host and a pathogen is complex and involves specific accommodations by both partners in the interaction. It is not possible to have a disease without a susceptible host, a pathogen, and a favorable environment. Fungi have adapted a number of strategies to infect plants and humans. The most comprehensive studies that involve the interaction of a host and fungal parasite have been conducted on plants. Fungal pathogens are heterotrophs capable of obtaining nutrients from both dead and living sources and can establish three types of trophic (e.g., feeding) relationships with plants: biotrophic, hemi-biotrophic, and necrotrophic. However, these relationships likely represent a continuum of interactions that occur been the host and fungal pathogen. In a biotrophic (obligate parasite) relationship, the fungus can live and multiply only on or within another living organism. Examples of obligate parasites (e.g., biotrophs) include fungi that cause mildew and rust diseases of plants. Interestingly, biotrophic relationships of fungi with animals and humans have not been documented. In a hemi-biotrophic relationship, the fungus is dependent on a living cell early in the infection process but then obtains nutrients from dead plant cells as a necrotroph. Examples of this include the smut and leaf curl diseases of plants. Many hemi-biotrophs have a yeast-like phase early in their life cycle and then convert to mycelial growth habit to cause infection of plants. In general, necrotrophy is the most common type of relationship between plants and fungi. In this type of relationship, the fungus obtains nutrients from dead cells often killed in advance by toxins or enzymes. Examples of necrotrophs include *Sclerotinia sclerotiorum*, a pathogen of more than 408 species of plants, *Bipolaris maydis*, the causal agent of Southern corn leaf blight, and species of *Cercospora*, pathogens of hundreds of plants.

Because host-pathogen interactions are complex, it is often difficult to identify specific factors required for pathogenicity. Consequently, only a few fungal traits have been shown conclusively to be involved in pathogenicity on either plants or humans. Traits identified as being required for pathogenicity include the ability of the fungus to produce adhesive compounds, melanins, toxins, and enzymes. Further, parasitic growth is associated with the transcription of a suite of genes, the expression of which is coordinated by elaborate signaling networks involved in sensing environmental stimuli and transmitting the molecular communication between the pathogen and the host.[32–38]

Whether a human or plant pathogen, the ability of the fungus to adhere to the host is of paramount important for pathogenicity. For example, plant pathogens secrete mucilage that allows spores to adhere to plant surfaces.[39] Likewise, the ability to adhere to host tissue and avoid expulsion during the early phase of infection has also been shown to be important for fungal pathogens of humans.[30,37] Another factor important in the early phase of infection by both human and plant pathogens is melanin. Melanin is a component of the cell wall of most pathogenic fungi, and in animal pathogens it protects the invading fungus from reactive oxygen species produced by the host.[30] Such a role for melanin in fungal interactions with plants has not been shown, but a similar role is likely because reactive oxygen species are often produced as a defense mechanism to microbial infection. A demonstrated role for melanin in plant pathogenic fungi is to provide structural integrity of specialized infection structures known as appressoria. These structures accumulate glycerol, which creates osmotic pressure required for direct penetration of plant tissue by the fungus.[40,41]

Many pathogenic fungi are thigmotrophic and able to recognize the surface of human and plant cells. Fungi can sense the topography of host cells and respond by altering their morphology, growth rates, and adhesion to cell surfaces.[30,42] Such recognition is known to be important for the bean rust pathogen *Uromyces appendiculatus*[41] that infects through stomata of the leaf. Bean leaf stomata are slightly raised (0.5 μm) above the leaf surface, and the fungus recognizes the difference in leaf surface topography to locate stomata and initiate infection. Many human pathogens are also thigmotrophic, and this trait may be important in human tissue invasion.[30]

Toxins produced by some plant pathogenic fungi are pathogenicity factors and are required for disease development. Other toxins have no known role in pathogenicity. Most of these compounds show toxicity to several species of plants and are known as nonselective toxins.[43] Cercosporin, a photosensitizing toxin produced by species of *Cercospora*,[44] is an example of a nonselective toxin. Another class of toxins, known as host-selective toxins, is toxic to a specific plant genotype within a species, and they often determine the host range of the fungus. Over 20 species of fungi are known to produce host-selective toxins.[45,46]

Fungal toxins have not been shown to be pathogenicity factors for human fungal pathogens. This is interesting, as members of the Eurotiomycetes, a class of fungi with recognized pathogens of humans, are known to produce a number of toxins. These include gliotoxin, a known immunosuppressant and inducer of animal programmed cell death, produced by *Aspergillus fumigatus*,[30] and aflatoxin, a carcinogen and immunosuppressant, produced by *A. flavus*.[14] Also, most species of *Aspergillus* that are pathogens of humans are known to produce a number of toxic compounds that are harmful if ingested (mycotoxins), but none has been shown to be involved in pathogenicity. Similarly, no role has been shown for any of the numerous ribotoxins (e.g., alpha-sarcin, restrictocin, and mitogillin) produced by filamentous fungi.[30]

Filamentous fungi are known to secrete a wide array of extracellular enzymes that are important for degradation of complex molecules during saprophytic growth. However, it has been difficult to demonstrate a requirement for these degradative enzymes in the pathogenicity of plants or animals.[47,48] Gene families often encode degradative enzymes, and creating mutant strains lacking all enzymatic activity is difficult. Further, some enzymes are induced only in host tissue or produced only during a certain phase of growth.[30]

Regardless of the interrelationship, the interaction of a fungus with its host involves molecular communication between the two, and many genes shown to regulate pathogenicity are components of signaling pathways. Fungi are known to have a number of signaling pathways including MAP kinases, G-proteins, and phosphokinases that have been shown to be required for pathogenicity.[32,38] Research in this field may potentially lead to disease management strategies based on the inhibition of important metabolic and regulatory pathways in fungi.

EPIDEMIOLOGY

Epidemiology involves the study of disease epidemics and factors that influence them over time. An epidemic is defined as a widespread and severe disease outbreak in a host population. An epidemic that occurs over a wide geographic region often for several years is referred to as a pandemic. Various disease epidemics caused by fungi or fungus-like organisms on plants have occurred throughout the world, including, chestnut blight (*Cryphonectria parasitica*), Dutch elm disease (*Ophiostoma novo-ulmi*), late blight Irish potato famine (*Phytophthora infestans*), southern corn leaf blight (*Bipolaris maydis*), and wheat rust (*Puccinia graminis* f.sp. *tritici*). Severe and widespread chestnut blight and Dutch elm disease epidemics were caused by plant pathogens introduced to geographic areas outside their naturalized range.

The examination of disease increase within a host population requires a thorough understanding of interactions that occur among the pathogen, host, and environment over time. In particular, an understanding of pathogen dispersal, ecology, population biology, reproduction, survival, and trophic behavior is of paramount importance for becoming knowledgeable with components of disease epidemiology. For example: (1) How dependent is the fungus on its host for survival and what are its mechanisms of survival in absence of a host? (2) How much inoculum (e.g., mycelium, spores, or sclerotia of a fungus that cause disease) is present at the beginning of an epidemic? (3) What propagule(s) of the fungus serves as initial (primary) inoculum for epidemic development? (4) Does the fungus produce additional inoculum (secondary inoculum) during the epidemic? (5) Is there one (monocyclic epidemic) or multiple (polycyclic epidemic) cycle(s) of infection per host cycle per year? and (6) What are the mechanisms for short- and long-distance dispersal of inoculum (e.g., animals, insects, humans, plant material, seed, soil, water, and wind)? The mechanisms of dispersal and amount of primary inoculum have a pronounced effect on the characteristics and scale of plant disease epidemics. Some propagules have limited dispersal ranging from millimeters to less than a meter, while others can be dispersed from hundreds to thousands of kilometers within and across continents.[49] The mode of dispersal also influences the pattern and spatial distribution of disease (i.e., random, regular, or aggregated) and subsequent development of epidemics. Monocyclic epidemics are very dependent on the initial amount of inoculum present at the beginning of the epidemic, because the fungus does not produce secondary inoculum for subsequent infection. In contrast, the amount of initial inoculum is less important for polycyclic epidemics, because there are usually several cycles of inoculum production and dispersal accompanied by repeated cycles of host infection and symptom development.

The assessment and measurement of plant disease are fairly straightforward and often expressed in terms of disease incidence (number or percentage of diseased individuals in a population) and disease severity (proportion of total area or the amount of plant tissue diseased). Epidemics are measured in terms of disease progress over a period of time. This is accomplished by plotting either disease incidence or disease severity against time. Various mathematical models (e.g., exponential, Gompertz, linear, logistic, monomolecular, Weibul, etc.) can be employed to calculate the rate of disease progress and area under the disease progress curve (AUDPC). Additional components of disease progress and epidemic development such as the incubation period (length of time from inoculation to appearance of disease symptoms) and latent period (length of time from inoculation to spore production) can also be estimated. For more information on plant disease epidemiology see Campbell and Madden[49] and the Arneson website: http://www.apsnet.org/education/AdvancedPlantPath/Topics/Epidemiology/Epidemiology.htm.

Plant diseases are usually not caused by a single individual, but rather a population of genetically distinct individuals. Therefore, the importance of understanding the genetic diversity and structure of pathogen populations should not be ignored when examining disease epidemics. Taken collectively with the multitude of factors that contribute to the development of plant disease epidemics, this information will provide a better understanding of epidemiological processes to prevent the spread of infectious disease-causing agents in the future.

In general, fungi causing invasive human diseases do not usually spread from one individual to another;[27] however, species of *Candida* may contaminate donor organs and medical instruments, including catheters and prosthetic devices.[50] *Candida* species have also been found on the hands of hospital workers,[50] although the significance of this in disease transmission is not known. Prevailing evidence shows that most fungal human pathogens are endemic in the environment and that inoculum is produced by the fungus growing as a saprophyte on a number of substrates. *Cryptococcus neoformans*, for example, has been isolated from pigeon droppings and Eucalyptus trees.[33] Fisher et al.[51] found that the major epidemic of coccidioidomycosis that occurred between 1991 and 1994 in California was from natural populations and not newly evolved or introduced pathogenic strains of *Coccidioides immitis*. They also observed that serious outbreaks of coccidioidomycosis often occurred after heavy rains following long periods of drought.[51] The source of *A. fumigatus* inoculum remains unresolved. It is unclear whether spores of the fungus originate from within the hospital or from an outside source.[50] There is also evidence that the drinking water supply can be contaminated with propagules of *A. fumigatus*.[50]

DETECTION AND IDENTIFICATION

The development of PCR and automated DNA sequencing techniques has contributed significantly to our ability to characterize and identify fungi. Numerous sequence-based methods are currently available for the detection and identification of fungi isolated in pure culture and from environmental and tissue samples. Most of these methods have focused on the utilization of DNA sequences that encode for the nuclear ribosomal RNA (rDNA) subunit. The degree of nucleotide sequence variation in this region allows for the differentiation of fungi belonging to different taxonomic groups, and is also useful for their diagnoses and identification (Figure 6.7).

A commonly employed strategy for diagnosis and identification of fungi is presented in Figure 6.8. The procedure involves sequencing several rDNA regions that provide different and complementary levels of phylogenetic resolution.[52,53] The transcribed 5S, 5.8S, 28S large subunit (LSU), and 18S

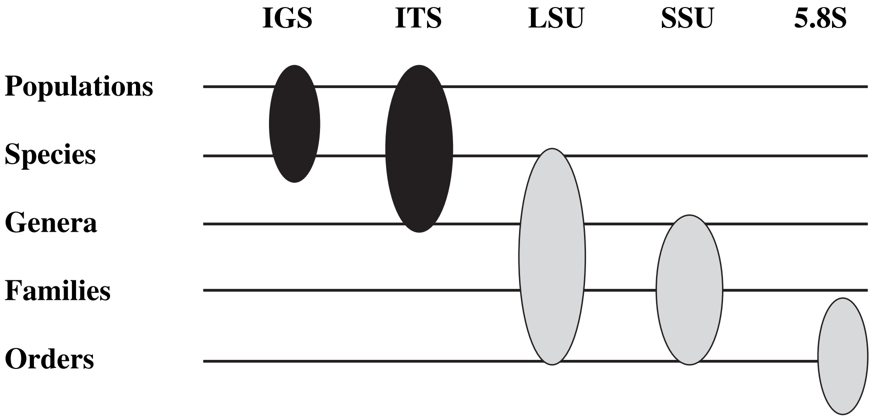


FIGURE 6.7 Levels of taxonomic resolution provided by nucleotide sequence data of nuclear ribosomal DNA (rDNA) genes. IGS, intergenic spacer region; ITS, internal transcribed spacer regions (ITS1 and ITS2); LSU, 28S large ribosomal subunit gene; 5.8S ribosomal subunit gene; and SSU, 18S small ribosomal subunit gene. Figure kindly provided by R. Vilgalys, Department of Biology, Duke University.

small subunit (SSU) rDNA regions usually provide resolution at or above the genus level, while the nontranscribed intergenic spacer (IGS) and internal transcribed spacer regions (ITS1 and 2) are more rapidly evolving and usually provide resolution at or below the species level. Initially the IGS and/or ITS regions are sequenced, and the Basic Local Alignment Search Tool (BLAST) search algorithm http://www.ncbi.nlm.nih.gov/ is utilized to determine whether there is a significant match (>70% identity) among sequences currently deposited in a database [GenBank (USA), European Molecular Biology Laboratory (EMBL) (Europe), and DNA Database of Japan (DDBJ) (Japan)]. If a match of greater than 70% is identified, the sequence is aligned with other similar sequences in the database and subjected to phylogenetic and distance analyses to further optimize sequence identification. Support for the link between the unknown sequence and its closest known sequence is evaluated using the additional statistical methods that provide an indication of their relatedness. In cases where a significant match is not identified, additional regions of the genome that provide resolution above the species level (e.g., SSU, LSU, 5.8S) can be sequenced, aligned, and subjected to phylogenetic analysis for sequence identification. By using this approach with known reference sequences and in conjunction with other mitochondrial ribosomal RNA (rDNA) and protein-encoding genes such as actin, ATPase, α- and β-tubulin, chitin synthase, hydrophobins, laccase, and translation elongation factor 1-α, it is possible to identify most fungi at various taxonomic levels.[23,52,54,55]

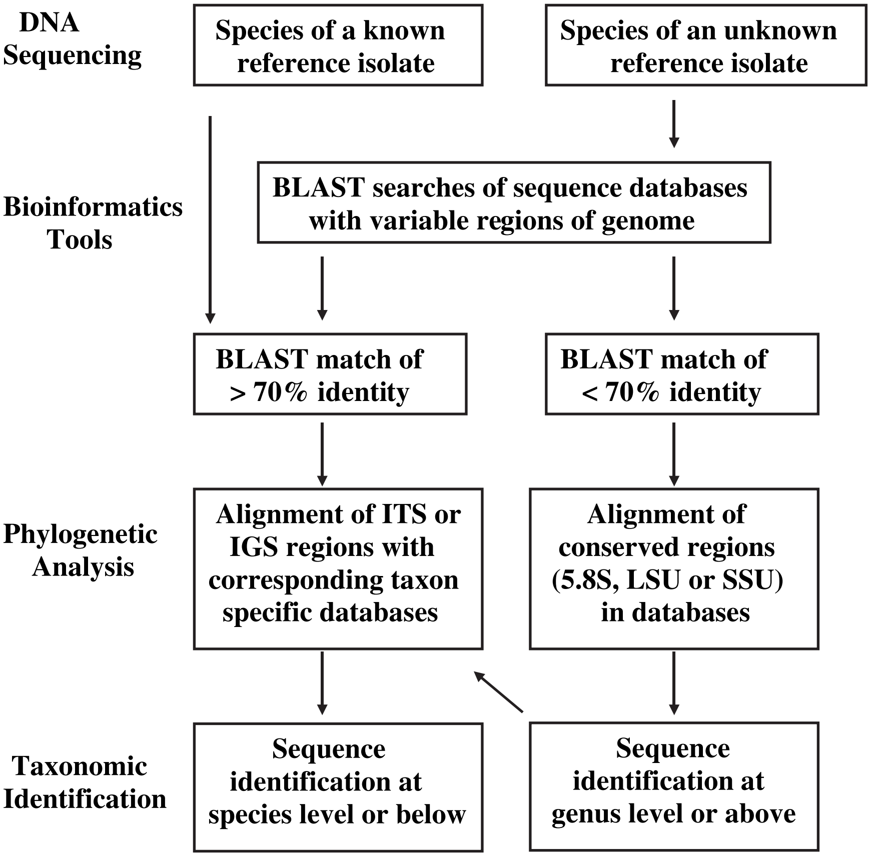

DNA Sequencing
Species of a known reference isolate
Species of an unknown reference isolate
Bioinformatics Tools
BLAST searches of sequence databases with variable regions of genome
BLAST match of > 70% identity
BLAST match of < 70% identity
Phylogenetic Analysis
Alignment of ITS or IGS regions with corresponding taxon specific databases
Alignment of conserved regions (5.8S, LSU or SSU) in databases
Taxonomic Identification
Sequence identification at species level or below
Sequence identification at genus level or above

FIGURE 6.8 Molecular-based identification of fungi with ribosomal RNA (rDNA) gene sequences. Figure kindly provided by R. Vilgalys, Department of Biology, Duke University).

Although databases have been demonstrated to be useful in identifying a wide range of fungi, in certain instances, these available databases may not be sufficient to place a DNA sequence within a previously sampled genus or species. The nontranscribed spacer regions of the rDNA subunit can be employed to identify genetically distinct individuals in a species at the population level. However, additional regions of the fungal genome are usually needed for this purpose. The detection of DNA sequence variation from multiple regions of the fungal genome coupled with appropriate statistical analysis of the data provides a very powerful tool for identification of fungi at or below the species level. This approach is referred to as multilocus sequence typing (MLST) and offers great promise for assigning unknown fungal individuals to a population of origin.[56]

In the past 15 years, mycologists and plant pathologists have employed DNA fingerprinting and PCR-based techniques to develop multilocus genetic markers for the identification of genetically distinct individuals in fungal pathogen populations. These population genetics studies have contributed significantly to our understanding of the genetic diversity and structure of populations of plant pathogens, and have provided a conceptual framework for defining and differentiating genetically distinct individuals (clones). Information on the genetic diversity and structure of fungal populations has resulted in the development of appropriate sampling protocols required to differentiate introduced and indigenous genetic individuals. As newer techniques (real-time PCR and microfluidic methods) and more mitochondrial and nuclear DNA sequence data become available, rapid identification of genetically distinct individuals in field populations will be possible, facilitating fungal forensics.

REFERENCES

1. Kirk PM, Cannon PF, David JC, Staplers JA. *Ainsworth and Bisby's Dictionary of the Fungi*. 9th ed. Wallingford: CAB International, 2001.
2. Alexopoulos CJ, Mims CW, Blackwell M. *Introductory Mycology*. 4th ed. New York: John Wiley and Sons, 1996.
3. Lutzoni F, Pagel M, Reeb V. Major fungal lineages are derived from lichen symbiotic ancestors. *Nature* 2001;411:937–940.
4. Wiebe MG. Myco-protein from *Fusarium venenatum*: A well-established product for human consumption. *Appl Microbiol Biotechnol* 2002;58:421–427.
5. Kennedy J, Auclair K, Kendrew SG, Park C, Vederas JC, Hutchinson CR. Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science* 1999;284:1368–1372.
6. Kimura M, Anzai H, Yamaguchi I. Microbial toxins in plant-pathogen interactions: Biosynthesis, resistance mechanisms, and significance. *J Gen Appl Microbiol* 2001;47:149–160.
7. Manzoni M, Rollini N. Biosynthesis and biotechnological production of statins by filamentous fungi and application of these cholesterol-lowering drugs. *Appl Microbiol Biotechnol* 2002;58:555–564.
8. Esser K, Lemke PA. The Mycota. *In* Osiewacz HD, ed. *Indust Appl*, Vol. 10. Berlin, 1994.
9. Puchart V, Katapodis P, Biely P, Kremnicky L, Christakopoulos P, Vrsanska M, Kekos D, Macris BJ, Bhat MK. Production of xylanases, mannanases, and pectinases by the thermophilic fungus *Thermomyces lanuginosus*. Enzyme Microb Technol 1999;24:355–361.
10. Swift RJ, Karandikar A, Griffen AM, Punt PJ, van den Hondel CA, Robson GD, Trinci AP, Wiebe MG. The effect of organic nitrogen sources on recombinant glucoamylase production by *Aspergillus niger* in chemostat culture. *Fungal Genetics and Biology* 2000;31:125–133.
11. Tsukagoshi N, Kobayashi T, Kato M. Regulation of the amylolytic and (hemi-)cellulolytic genes in aspergilli. *J Gen Appl Microbio* 2001;47:1–19.

12. Farr DF, Bills GF, Chamuris GP, Rossman AY. *Fungi on Plants and Plant Products in the United States*. St. Paul: APS Press, 1989.
13. Larone DH. *Medically Important Fungi: A Guide to Identification*. 4th ed. Washington, D.C.: ASM Press, 2002.
14. Richard JL, Payne GA. Mycotoxins: Risks in Plant, Animal and Human Systems. Council for Agricultural Science and Technology 2003.
15. Benjamin DR. *Mushrooms Poisons and Panaceas: A Handbook for Naturalists, Mycologists, and Physicians*. New York: W.H. Freeman and Company, 1995.
16. Whittaker RH. New concepts of kingdoms of organisms. *Science* 1969;163:150–160.
17. Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 2000;290:972–977.
18. Galagan JE, Calvo SE, Borkovich KA, et al. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 2003;422:859–868.
19. Griffin DH. *Fungal Physiology*. 2nd ed. New York: John Wiley and Sons, 1994.
20. Dick MW. *Straminipilous Fungi*. Dordrecht: Kluwer Academic Publishers, 2001.
21. Madigan MT, Martinko JM, Parker J. *Brock Biology of Microorganisms*. 10th ed. Upper Saddle River: Prentice Hall, 2003.
22. Redecker D. New views on fungal evolution based on DNA markers and the fossil record. *Research in Microbiology* 2002;153:125–130.
23. James TY, Porter D, Leander CA, Vilgalys R, Longcore JE. Molecular phylogenetics of the Chytridiomycota supports the utility of ultrastructural data in chytrid systematics. *Canad J Botany* 2000;78:336–350.
24. Schubler A, Schwarzott D, Walker C. A new phylum, the Glomeromycota: phylogeny and evolution. *Mycol Res* 2001;105:1413–1421.
25. Berbee ML, Taylor JW. Dating the evolutionary radiations of the true fungi. *Canad J Botany* 1993;71:1114–1127.
26. Agrios GN. *Plant Pathology*. 4th ed. San Diego: Academic Press, 1997.
27. Berbee ML. The phylogeny of plant and animal pathogens in the Ascomycota. *Physiol Mol Plant Pathol* 2001;59:165–187.
28. Patel R, Paya CV. Infections in solid-organ transplant recipients. *Clin Microbiol Rev* 1997;10:86–124.
29. Ribes JA, Vanover-Sams CL, Baker DJ. Zygomycetes in human disease. *Clin Microbiol Rev* 2000;13(2):236–301.
30. van Burik JAH, Magee PT. Aspects of fungal pathogenesis in humans. *Ann Rev Microbiol* 2001;55:743–772.
31. Ellis M. Invasive fungal infections: evolving challenges for diagnosis and therapeutics. *Molr Immunol* 2002;38:947–957.
32. D'Souza CA, Heitman J. Conserved cAMP signaling cascades regulate fungal development and virulence. *FEMS Microbiol Rev* 2001;25:349–364.
33. Hull CM, Heitman J. Genetics of *Cryptococcus neoformans*. *Ann Rev Genetics* 2002;36:557–615.
34. Kahmann R, Basse C. Fungal gene expression during pathogenesis-related development and host plant colonization. *Curr Op Microbiol* 2001;4:374–380.
35. Kronstad J, De Maria A, Funnell D, Laidlaw DR, Lee N, Moniz de Sa' M, Ramesh M. Signaling via cAMP in fungi: interconnections with mitogen-activated protein kinase pathways. *Arch Microbiol* 1998;170:395–404.
36. Xu JR. MAP Kinases in fungal pathogens. *Fungal Genetics Biol* 2000;31:137–152.
37. Alonso-Monge R, Navarro-Garcia F, Roman E, Eisman B, Nombela C, Pla J. Strategies for the identification of virulence determinants in human pathogenic fungi. *Curr Genetics* 2003; 42:301–312.

38. Hube B. From commensal to pathogen: stage- and tissue-specific gene expression of *Candida albicans*. *Curr Op Microbiol* 2004;7:336–341.
39. Idnurm A, Howlett BJ. Pathogenicity genes and phytopathogenic fungi. *Mol Plant Pathol* 2001;2:241–255.
40. DeJong JC, McCormack BJ, Smirnoff N, Talbot NJ. Glycerol generates turgor in rice blast. *Nature* 1997;389:244–245.
41. Deising HB, Werner S, Wernitz M. The role of fungal appressoria in plant infection. *Microbes Infection* 2000;2:1631–1641.
42. Jaffe MJ, Leopold AC, Staples RC. Thigmo responses in plants and fungi. *American J Botany* 2002;89:375–382.
43. Ballio A. Non-host-selective fungal phytotoxins–biochemical aspects of their mode of action. *Experientia* 1991;47:783–790.
44. Daub ME, Ehrenshaft M. The photoactivated Cercospora toxin cercosporin: Contributions to plant disease and fundamental biology. *Ann Rev Phytopathol* 2000;38:461–490.
45. Markham J, Hille J. Host-selective toxins as agents of cell death in plant-pathogen interactions. *Mol Plant Pathol* 2001;2:229–239.
46. Wolpert TJ, Dunkle LD, Ciuffetti LM. Host-selective toxins and avirulence determinants: What's in a name? *Ann Rev Phytopathol* 2002;40:251–285.
47. Isshiki A, Akimitsu K, Yamamoto M, Yamamoto H. Endopolygalacturonase is essential for citrus black rot caused by *Alternaria citri* but not brown spot caused by *Alternaria alternata*. *Mol Plant-Microbe Interactions* 2001;14:749–757.
48. Kim H, Ahn JH, Gorlach JM, Caprari C, Scott-Craig JS, Walton JD. Mutational analysis of beta-glucanase genes from the plant-pathogenic fungus *Cochliobolus carbonum*. *Mol Plant-Microbe Interactions* 2001;14:1436–1443.
49. Campbell CL, Madden LV. *Introduction to Plant Disease Epidemiology.* New York: North Carolina State University Document, 1990.
50. Kullberg BJ, Lashof A. Epidemiology of opportunistic invasive mycoses. *Eur J Med Res* 2002;7:183–191.
51. Fisher MC, Koenig GL, White TJ, Taylor JW. Pathogenic clones versus environmentally driven population increase: Analysis of an epidemic of the human fungal pathogen *Coccidioides immitis*. *J Clin Microbiol* 2000;38:807–813.
52. Binder M, Hibbett DS. Higher-level phylogenetic relationships of homobasidiomycetes (mushroom-forming fungi) inferred from four rDNA regions. *Mol Phylogenetics Evol* 2002;22:76–90.
53. Bruns TD, Szaro TM, Gardes M, Cullings KW, Pan JJ, Taylor DL, Horton TR, Kretzer A, Garboletto M, Li Y. A sequence database for the identification of ectomycorrhizal Basidiomycetes by phylogenetic analysis. *Mol Ecol* 1998;7:257–272.
54. Gardes M, Bruns TD. ITS primers with enhanced specificity for basidiomycetes–application to the identification of Mycorrhizae and rusts. *Mol Ecol* 1993;2:113–118.
55. Carbone I, Kohn LM. Multilocus nested haplotype networks extended with DNA fingerprints show common origin and fine-scale, ongoing genetic divergence in a wild microbial metapopulation. *Mol Ecol* 2001;10:2409–2422.
56. Taylor JW, Fisher MC. Fungal multilocus sequence typing–It's not just for bacteria. *Curr Op Microbiol* 2003;6:1–6.

CHAPTER 7

Forensic Aspects of Biologic Toxins

JAMES D. MARKS
Department of Anesthesia and Pharmaceutical Chemistry, University of California, San Francisco, San Franciso General Hospital, San Francisco, California

INTRODUCTION

Biologic toxins are molecules produced by living organisms that are poisonous to other species, such as humans. Some biologic toxins are so potent and relatively easy to produce that they have been classified as biothreat agents. These include the botulinum neurotoxins, classified as Category A agents, and ricin, staphylococcal enterotoxin B, and *Clostridia perfringens* epsilon toxin, all classified as Category B agents (see http://www.niaid.nih.gov/biodefense/bandc_priority.htmj for classification of biothreat agents). These four biothreat agents are all proteins composed of amino acid building blocks. As such, they have a number of features that distinguish them from viral or bacterial threat agents. First, they are not contagious, as the threat agent is not a living organism. For the same reason, these agents cannot be routinely cultured from either patients or the environment after exposure, making forensic detection more difficult. Since proteins are composed of amino acids and not nucleic acid, it is also not possible to amplify and detect the presence of toxins using polymerase chain reaction (PCR), or by any type of classic DNA hybridization technology. Rather, detection typically relies on the use of antibodies, to bind to and detect the presence of toxins, or newer detection technologies such as mass spectrometry.

This chapter will focus on the four biothreat toxins described above, and their forensic aspects. The majority of the chapter will be spent on the botulinum neurotoxins, as these are the most poisonous substances known,

This work was partially supported by Department of Defense Contract DAMD-17-98-C-8030 and NIAID grants AI056493 and AI53389-01

Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

and the only toxins classified as Category A threat agents. The remainder of the chapter will be devoted to sections on the other three biothreat toxins: ricin, staphylococcal enterotoxin B, and *Clostridia perfringens* epsilon toxin.

BOTULINUM NEUROTOXIN AND BOTULISM

Botulism is a rare but life-threatening disease caused by spore-forming bacteria of the *Clostridia* genus, including *Clostridium botulinum*, *C. baratii,* and *C. butyricum*[1]. The disease results from bacterial secretion of botulinum neurotoxin, the most poisonous substance known.[2] Approximately 7 pg of pure neurotoxin is the LD50 for a mouse, and it has been estimated that the human LD 50 is approximately 0.09–0.15 μg intravenously, 0.7–0.9 μg inhalationally, and 70 μg orally.[3–6]

Botulism is characterized by prolonged paralysis, which if not immediately fatal requires prolonged hospitalization in an intensive care unit (ICU) and mechanical ventilation. The potent paralytic abilities of the neurotoxin have also resulted in its development as a biowarfare and biothreat agent,[7] as well as a medicine to treat a range of overactive muscle conditions including cervical dystonias, cerebral palsy, post-traumatic brain injury, and post-stroke spasticity.[8]

Clostridial neurotoxins differ significantly from each other in their amino acid sequence, resulting in the elicitation of different antibody responses. The different antibody responses allow the classification of the neurotoxins into different serotypes; antibodies that recognize one serotype do not recognize other serotypes. There are seven neurotoxin serotypes (A, B, C, D, E, F, and G),[9,10] four of which (A, B, E, and F) cause naturally occurring human botulism.[7]

Naturally occurring botulism can result from ingestion of preformed toxin (food botulism) or from toxin produced *in situ* due to wound infection (wound botulism) or colonization of the gastrointestinal tract (infant or intestinal botulism). Botulism can also occur in exposed laboratory workers or from an overdose of therapeutic neurotoxin. In addition, the botulinum neurotoxins are classified by the Centers for Disease Control (CDC) as one of the six highest-risk threat agents for bioterrorism, due to their extreme potency and lethality, ease of production and transport, and victims' need for prolonged intensive care.[7] Intoxication can occur via oral ingestion of toxin or inhalation of aerosolized toxin.[11,12] While only four of the neurotoxin serotypes cause natural human disease, aerosolized neurotoxin serotypes C, D, and G produce botulism in primates by the inhalation route,[11] and would most likely also affect humans. Thus it is likely that any one of the seven neurotoxin serotypes can be used as a biothreat agent. Due to the severity of illness and the potential for outbreaks, both foodborne and biothreat botulism are public health emergencies.

TYPES OF BOTULISM

Five types of botulism occur in humans: foodborne, wound, infant, intestinal, and inadvertent. A sixth type, intentional or bioterror botulism, is likely to occur during our lifetimes. Each type is associated with different epidemiology and pathogenetic mechanisms. The first recognized case of botulism in the U.S. occurred in 1899 and was caused by a beef tamale.[13] Food botulism was the most common form of botulism in the U.S. prior to 1980.[1] Infant (or intestinal) botulism was first described in 1976 by two groups[14,15] and is now the most frequently reported type of botulism in the U.S.[1] Wound botulism was first described in the U.S. in 1951, with initial cases primarily due to traumatic wounds of the extremities.[16] More recently, the incidence of this form of botulism has increased and has been associated with injection drug users injecting black tar heroin.[17] An adult variant of infant botulism, variously called botulinal autointoxication, or hidden, adult intestinal, or adult infectious botulism was first described in 1979.[18–20] Inadvertent botulism results from unintentional exposure and typically occurs in laboratory workers[21] and in patients receiving therapeutic botulinum neurotoxin.[22]

INTENTIONAL BOTULISM

While successful use of neurotoxin as a bioterror agent has not occurred, it is likely only a matter of time until botulism is intentionally caused by release of toxin by terrorists. The toxin's potency and lethality make it an ideal bioweapon, and has resulted in its classification by the CDC as a category A biothreat agent, the highest level. Botulinum toxin has already been released unsuccessfully by the Japanese cult Aum Shinryko.[7] Both Iraq and the former Soviet Union produced botulinum neurotoxin (BoNT) for use as weapons[23,24] and at least three additional countries (Iran, North Korea, and Syria) have developed or are believed to be developing BoNT as an instrument of mass destruction. Iraq produced 19,000 L of concentrated BoNT, of which 10,000 L were weaponized in missile warheads or bombs.[23,25]

Exposure of even a small number of civilians to botulinum neurotoxin would overwhelm the health care delivery system of any metropolitan center. Treatment of botulism requires prolonged ICU hospitalization and mechanical ventilation for up to 6 weeks. With the downsizing and closing of hospitals, most ICUs run at 80%–100% occupancy. In San Francisco, for example, there are approximately 210 ICU beds, with an average occupancy rate of greater than 90%. As few as 30 cases of botulism would fill all empty ICU beds and occupy them for up to 6 weeks. This would eliminate availability of ICU beds for postoperative patients requiring ICU care, such as organ transplantation, neurosurgery, cardiac surgery, and traumatic injuries. Patients requiring such operations would represent "collateral damage," with necessary surgery postponed, or transferred to outlying hospitals. Major civilian exposure to BoNT would have catastrophic effects. One study estimated that aerosol exposure of 100,000 individuals to toxin, as could occur with an aerosol release over a metropolitan area, would result in 50,000 cases with 30,000 fatalities.[26] Such exposure would result in 4.2 million hospital days and an estimated cost of \$8.6 billion. In this study, the most important factors reducing mortality and cost were early availability of antitoxin and mechanical ventilation.[15] Such treatment could reduce deaths by 25,000 and costs by \$8.0 billion.

CLOSTRIDIAL BACTERIOLOGY

Botulism is caused by the actions of a 150-kDa neurotoxin secreted by spore-forming anaerobic bacteria of the genus *Clostridia*. *C. botulinum* can be classified into at least four genetically and phenotypically diverse groups (I–IV).[27,28] While these groups are different enough to be classified as separate species, they have all been classified as *C. botulinum* since they share the common feature of neurotoxin production. The organisms in group I are referred to as proteolytic, and the organisms in group II as nonproteolytic, based on their ability to digest complex proteins. All serotype A strains are group I, serotype B and F can be produced by either group, and serotype E is produced by group II strains. Two additional *Clostridial* species, *C. butyricum* and *C. baratii*, have been found to produce neurotoxins E[29,30] and F,[31,32] respectively. Serotypes C and D are both produced by group III organisms. Type C is found in avian species, occurring in domestic flocks and massive outbreaks in wild waterfowl.[33–35] Type C also occurs in other animals such as dogs, mink, and cattle. Type D outbreaks are rare and associated with cattle.[36] A single human outbreak of type C and of type D food botulism have been reported.[37,38] Group IV was created to accommodate an organism isolated from a soil sample in Argentina that produces a unique neurotoxin (type G) that causes a flaccid paralysis in mice.[39,40] No clinical cases of type G botulism have been reported, though it has been isolated from autopsy specimens.[41] Recently, the name *C. argentinense* has been proposed for Group IV *Clostridia*.[42] Finally, rare strains of *Clostridia* have been reported that cause clinical disease and secrete more than one toxin, for example A and B (Ab), A and F (Af), B and F (Bf), and B and A (Ba).[28,43–46]

NEUROTOXIN STRUCTURE AND FUNCTION

Botulinum neurotoxins are secreted from *Clostridial* species as a protein complex with an apparent size of approximately 900 kDa.[47,48] This complex consists of the neurotoxin and a number of proteins collectively called neurotoxin-associated proteins (NAPs). The NAPs include proteins classified as hemagglutinins,[49,50] due to their ability to agglutinate red blood cells, and other proteins termed nontoxin nonhemagglutinins (NTNH).[51,52] These other proteins stabilize the toxin and protect it from environmental degradation during passage through the gastrointestinal tract.[53,54]

The protein neurotoxin is secreted as a single polypeptide chain of approximately 150 kDa that is nicked by proteases to form a 100-kDa heavy chain and a 50-kDa light chain connected by a single disulfide bond. The sequences of the genes encoding neurotoxin serotypes A,[55–57] B,[58,59] C,[60] D,[61] E,[62] F,[52,63] and G[64] have been determined. While these toxins differ by as much as 65% at the amino acid level, it is likely that they share the same general protein fold.[10] Significant sequence variability has also been observed within toxin serotypes,[55,57–59] for example A1 and A2 types of neurotoxin A have been reported, which differ by 10% at the amino acid level.[55,57] The differences are highest in the heavy chain. Such differences may account for the reported differences in the ability of monoclonal antibodies to recognize toxins from different A strains.[65,66]

Recently, the X-ray crystal structure of types A and B neurotoxins have been solved at high resolution (Figure 7.1).[67–69] The structural studies, combined with functional studies, provide clear insight into how the botulinum neurotoxins interfere with normal release of the neurotransmitter acetylcholine, resulting in flaccid paralysis (Figure 7.2). The C-terminal portion of the heavy chain (H_C) comprises the receptor binding domain, which binds to cellular receptors on presynaptic neurons, resulting in toxin endocytosis[70,71] (Figure 7.3). The precise determination of the cellular receptors is unknown, but the presence of two coreceptors, a protein and a sialoganglioside such as G_{D1b} or G_{T1b}, have been proposed.[71–73] The binding domain consists structurally of an N-terminal subdomain consisting of a jelly roll motif and a C-terminal subdomain consisting of a β-trefoil motif. It is hypothesized that the C-terminal domain comprises the ganglioside binding site.[67] The N-terminal portion of the heavy chain (H_N) (Figure 7.1) comprises the translocation domain, which consists of α-helices and is involved in pore formation. It is hypothesized that the lower pH of the endosome induces a conformational change in this domain that creates a pore allowing the light chain to escape the endosome (Figure 7.3). The light chain (Figure 7.1) is a zinc endopeptidase which, depending on serotype, cleaves different members of the soluble -N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) family of proteins,

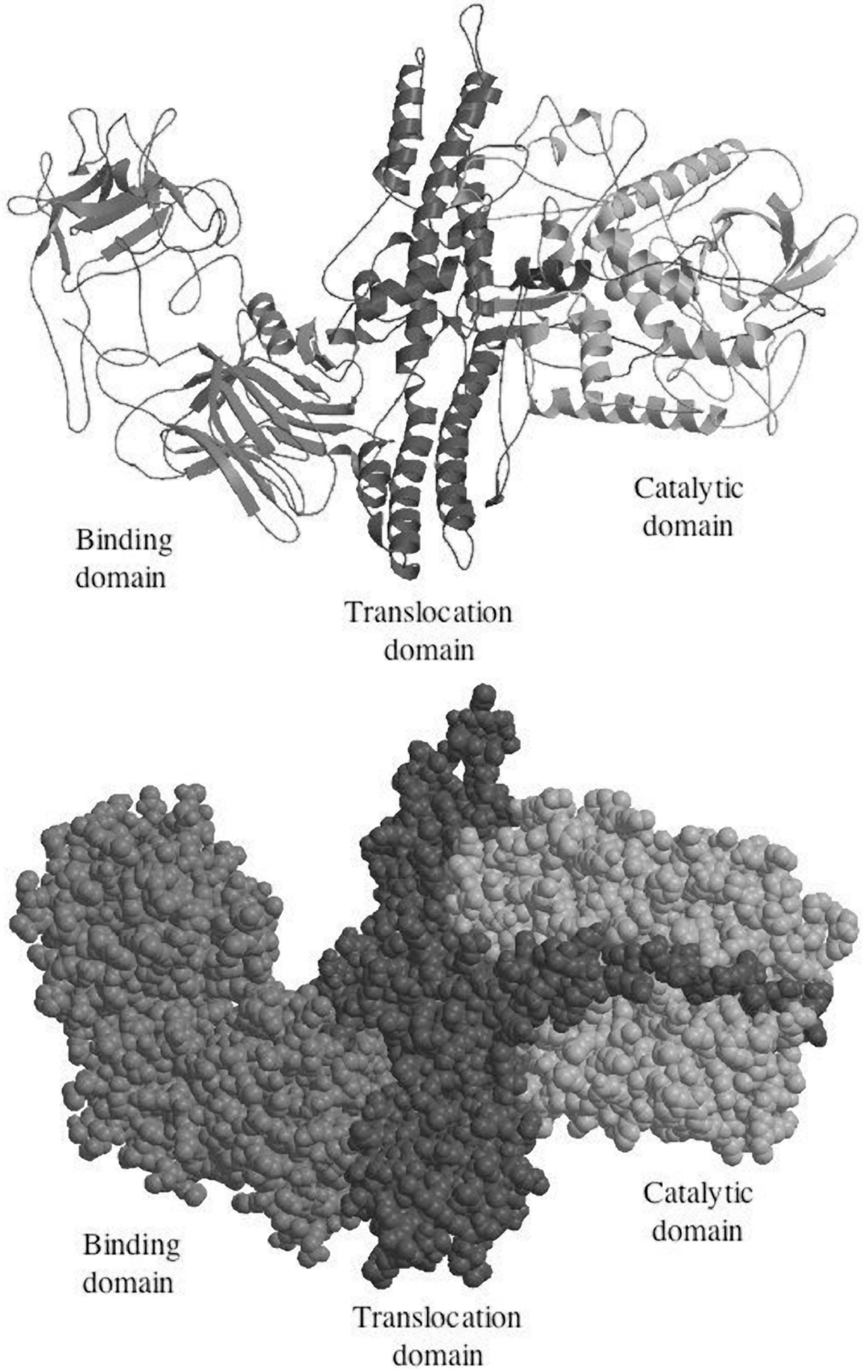

Binding domain
Translocation domain
Catalytic domain
Binding domain
Translocation domain
Catalytic domain

FIGURE 7.1 Atomic structure of botulinum neurotoxin type A. Ribbon (1A) and space filling (1B) models of the X-ray crystal structure of botulinum neurotoxin type A (ref. 64). The toxin consists of a binding domain, translocation domain, and catalytic domain, as described in detail in the text.

resulting in blockade of neuromuscular transmission[74,75] (Figures 7.2 and 7.3). The SNAREs are essential for normal fusion of the synaptic vesicle and acetylcholine release (Figure 7.1). Toxin serotypes A and E cleave distinct sites within SNAP-25 (synaptosomal-associated protein of 25 kDa);[75–78] serotypes B, D, F, and G cleave distinct sites within vesicle associated membrane protein (VAMP, also known as synaptobrevin);[74,75,79–82] and serotype C cleaves syntaxin and SNAP-25 (Figure 7.3).[83,84] These three SNARE proteins (syntaxin, SNAP-25, and synaptobrevin) interact to form a four-helix coiled-coil in a step that precedes synaptic fusion[85] (Figure 7.2). Cleavage of any one of these proteins blocks fusion and acetylcholine release, leading to a flaccid paralysis.

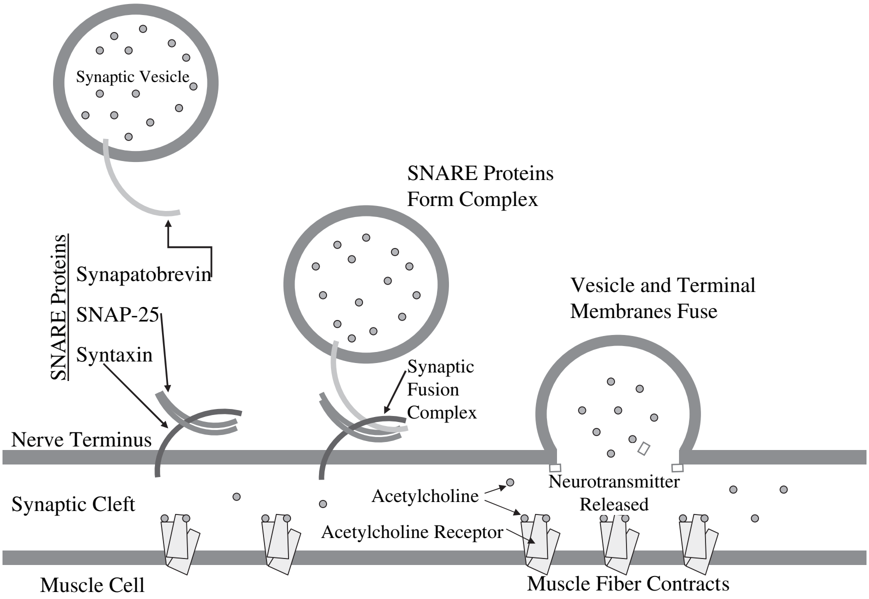

Synaptic Vesicle
SNARE Proteins
Synapatobrevin
SNAP-25
Syntaxin
SNARE Proteins Form Complex
Synaptic Fusion Complex
Vesicle and Terminal Membranes Fuse
Nerve Terminus
Synaptic Cleft
Acetylcholine
Neurotransmitter Released
Acetylcholine Receptor
Muscle Cell
Muscle Fiber Contracts

FIGURE 7.2 Normal neuromuscular transmission. Synaptic vesicles containing acetylcholine have the soluble -N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) synaptobrevin on their surfaces. Vesicular synaptobrevin interacts with the SNARE proteins syntaxin and synaptosomal-associated protein of 25 kDa (SNAP-25) to form a four-helix coiled-coil resulting in fusion of the synaptic vesicle with the presynaptic membrane. Acetylcholine is released from the vesicle, diffuses across the synaptic cleft, and binds to the acetylcholine receptor, resulting in normal muscle contraction.

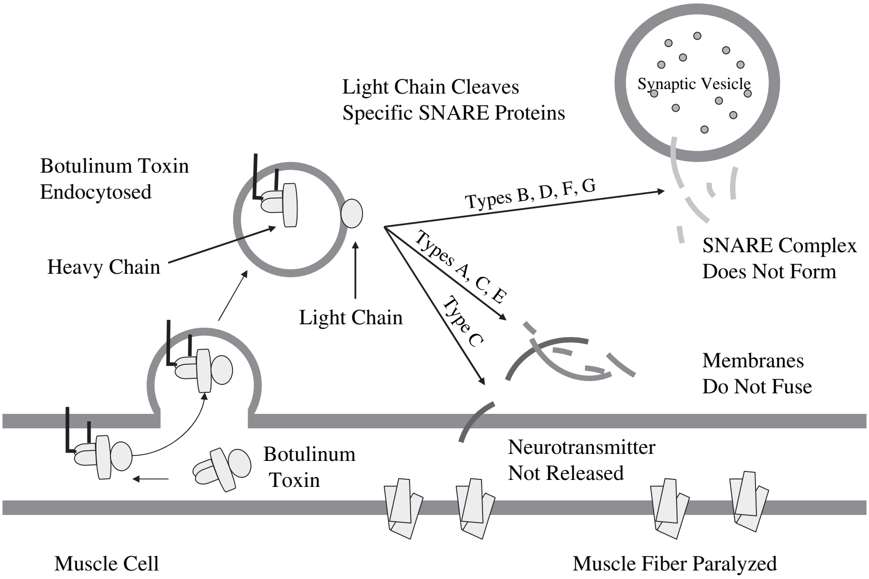

Light Chain Cleaves
Specific SNARE Proteins
Synaptic Vesicle
Botulinum Toxin
Endocytosed
Types B, D, F, G
Heavy Chain
Types A, C, E
SNARE Complex
Does Not Form
Type C
Light Chain
Membranes
Do Not Fuse
Botulinum
Toxin
Neurotransmitter
Not Released
Muscle Cell
Muscle Fiber Paralyzed

FIGURE 7.3 Effect of botulinum neurotoxin on normal neuromuscular transmission. Botulinum neurotoxin binds to unknown receptors on the presynaptic neuron membrane, resulting in endocytosis of the toxin. After endocytosis, the translocation domain changes conformation, resulting in release of the catalytic domain into the cytosol. Depending on the toxin serotype, the catalytic domain cleaves one or more members of the SNARE protein family. SNARE cleavage prevents formation of the SNARE complex and fusion of the vesicle with the membrane. As a result, acetylcholine is not released.

CLINICAL HISTORY, SYMPTOMS, AND FINDINGS OF BOTULISM

It is quite likely that the first indication of a bioterror attack with botulinum toxin will be the development of clinical disease (botulism) in exposed individuals. It is thus essential that forensic personnel be familiar with the symptoms of the disease and its differential diagnosis (other diseases that may be confused with botulism). The diagnosis of botulism is made clinically, with laboratory findings and confirmation not usually immediately available. The clinical syndrome of botulism is dominated by neurologic signs and symptoms resulting from blockade of neurotransmission at voluntary motor and cholinergic junctions.[86–88] Patients with botulism usually present with acute onset of weakness in muscles innervated by the cranial nerves, leading to diplopia, dysphonia, dysphagia, and dysarthria (Table 7.1). In mild cases, no other symptoms may develop. In more severe cases, symmetric weakness progresses in a

TABLE 7.1 Symptoms and signs of foodborne botulism. The data are from outbreaks of type A and B foodborne botulism reported in the U.S. in 1973–1974. The number of patients with available data varied from 35 to 55[120]

Symptoms	% Cases
Fatigue	77
Dizziness	51
Double vision	91
Blurred vision	65
Dysphagia	96
Dry mouth	93
Dysarthria	84
Sore throat	54
Dyspnea	60
Constipation	73
Nausea	64
Vomiting	59
Abdominal cramps	42
Diarrhea	19
Arm weakness	73
Leg weakness	69
Paresthesia	14
Signs	
Alert mental status	90
Ptosis	73
Gaze paralysis	65
Pupils dilated or fixed	44
Nystagmus	22
Facial palsy	63
Diminished gag reflex	65
Tongue weakness	58
Arm weakness	75
Leg weakness	69
Hyporeflexia or areflexia	40
Ataxia	17

descending manner, leading frequently to paralysis. If the illness is severe enough, the respiratory muscles are involved, leading to ventilatory failure and death unless intubation and mechanical ventilation are instituted. Intubation was required in 67% of type A botulism, 52% of type B botulism, and 39% of type E botulism.[89] Patients may also have evidence of autonomic dysfunction including dry mouth, blurred vision, orthostatic hypotension, urinary retention, and constipation. Sensory abnormalities are usually absent, since only motor and autonomic nerves are affected. Similarly, mental function is usually not affected.

Paralysis from botulism can be quite long-lasting. Mechanical ventilation may be required for 2 to 8 weeks with foodborne botulism, with paralysis lasting as long as 7 months reported.[88] Symptoms of cranial nerve dysfunction and mild autonomic dysfunction may persist for more than a year.[90–92] In infants, hospital stay averages 1 month, with serotype A causing longer lasting disease (5.4-week average hospitalization) than serotype B (3.8-week average hospitalization).[93] There is experimental evidence that neurotoxin catalytic activity persists at the nerve terminal, especially for serotypes A, B, and C, and that recovery initially results from the sprouting of new neuromuscular connections.[94,95]

EPIDEMIOLOGY OF INTENTIONAL BOTULISM AND DIFFERENTIAL DIAGNOSIS

The intentional release of botulism is most likely to be associated with the outbreak of a large number of cases of flaccid paralysis with prominent bulbar palsies. Other features may include an outbreak with an unusual toxin type (C, D, F, or G, vide infra), an outbreak with common geographic features but without a common dietary exposure, or multiple simultaneous outbreaks with no common source. The incubation period for intentional botulism is unknown, but is likely related to the route of exposure and amount of toxin administered. For oral exposure to toxin, one may obtain some idea of the incubation period from the foodborne botulism literature. Foodborne botulism has an incubation period of 6 hours to 10 days,[96] with the majority of cases developing between 18 to 72 hours after ingestion of contaminated food. It is difficult to know precisely the incubation period for aerosol exposure to botulinum toxin, due to the paucity of data. In one study, monkeys exhibited signs of intoxication 12 to 80 hours after aerosol exposure, with 4–7 monkey LD50s.[4] The incubation period for the three known cases of human botulism via the inhalation route was 72 hours.[21]

With respect to the source of toxin, waterborne delivery is unlikely, due to rapid inactivation by standard water treatments and the large inoculum of toxin required.[7] No cases of waterborne botulism have been reported. Botulinum toxin is stable for days in untreated water and beverages, however, which could be suitable vehicles for intentional toxin delivery.[7] Alternatively, toxin could be delivered via "nature's way" in food or food products, making it difficult to distinguish intentional toxin release from an outbreak of foodborne botulism.

Intentional or bioterror botulism must be distinguished from natural occurring botulism and from the many diseases that may mimic botulism. Recognition of botulism depends on astute clinicians who first see intoxicated patients. Unfortunately, naturally occurring botulism is underdiagnosed and frequently misdiagnosed, usually as a polyradiculopathy (Guillain Barré or Miller-Fisher syndrome), myasthenia gravis, or a disease of the central nervous system (Table 7.2). Common and uncommon misdiagnoses are listed in Table 7.2, along with features that distinguish botulism from these diseases. Botulism differs from other flaccid paralyses in: (1) its prominent cranial nerve palsies disproportionate to weakness below the neck; (2) the symmetry of the weakness; and (3) the absence of sensory changes, although approximately 14% of patients report paresthesias (Table 7.1).

Both intentional and naturally occurring botulism are much more likely to be associated with outbreaks (cluster of cases) than other diseases with which they may be confused such as Guillain Barré, poliomyelitis, or intoxications. Due to the mobility of populations, and a potentially long and variable incubation time, these cases may be separated both in time and space. This fact emphasizes the importance of prompt reporting of suspected botulism cases to the public health department by first-responders. In the early stages, foodborne botulism is most likely to be confused with intentional botulism, and in fact food could be a viable route for the intentional delivery of toxin. Foodborne botulism usually occurs in outbreaks where multiple individuals ingest contaminated food. From 1899 to 1996, 921 outbreaks of foodborne botulism were reported to the CDC, with a relatively constant incidence of approximately 9.5 outbreaks/year, with an average of 2.5 cases per outbreak.[1] The largest number of cases in a single foodborne outbreak was 59. Foodborne botulism is usually associated with ingestion of home canned products, most frequently foods low in acid such as vegetables, fish, or marine mammals, condiments, and meat products. Fruits are rarely involved, due to their high natural acidity. Outbreaks have also been reported for commercially prepared products[97,98] and for food prepared in restaurants.[99,100] In the U.S., the incidence of foodborne botulism is highest in Alaska, where the vehicle is typically native Alaskan foods consisting of fermented or salted fish or marine mammal products.

Foodborne botulism is typically caused by type A, B, and E neurotoxins. Of the 1,087 cases of foodborne botulism reported between 1950 and 1996, the

TABLE 7.2 Differential diagnosis of botulism. Common and infrequent misdiagnoses in patients with botulism (adapted from ref. 7)

Condition	Features that distinguish condition from botulism
Common misdiagnoses	
Guillain-Barré syndrome[a] and its variants, including Miller-Fisher syndrome	History of antecedent infection; paresthesias[b]; often ascending paralysis; early areflexia; eventual CSF protein increase; EMG findings
Myasthenia gravis[a]	Recurrent paralysis; EMG findings; sustained response to anticholinesterase therapy
Stroke[a]	Paralysis often asymmetric; abnormal CT or MRI scan
Intoxication with depressants (e.g. acute ethanol intoxication), organophosphates, carbon monoxide, or nerve gas	History of exposure; excessive drug levels detected in body fluids
Eaton-Lambert syndrome	Increased strength with sustained contraction; evidence of lung carcinoma; EMG findings similar to botulism
Tick paralysis	Paresthesias[b]; ascending paralysis; tick attached to skin
Other misdiagnoses	
Poliomyelitis	Antecedent febrile illness; asymmetric paralysis; CSF pleocytosis
CNS infections, especially of the brainstem	Mental status changes; CSF and EEG abnormalities
CNS tumor	Paralysis often asymmetric; abnormal CT or MRI scan
Streptococcal pharyngitis (pharyngeal erythema can occur in botulism	Absence of bulbar palsies; positive rapid antigen test result or throat culture
Psychiatric illness[a]	Normal EMG in conversion paralysis
Viral syndrome[a]	Absence of bulbar palsies and flaccid paralysis
Inflammatory myopathy[a]	Elevated creatine kinase levels
Diabetic complications[a]	Sensory neuropathy; few cranial nerve palsies
Hyperemesis gravidarum[a]	Absence of bulbar palsies and acute flaccid paralysis
Hypothyroidism[a]	Abnormal thyroid function test results
Laryngeal trauma[a]	Absence of flaccid paralysis; dysphonia without bulbar palsies
Overexertion[a]	Absence of bulbar palsies and acute flaccid paralysis

CSF, cerebrospinal fluid; EMG, electromyogram; CNS, central nervous system; EEG, electroencephalogram; CT, computed tomography; MRI, magnetic resonance imaging.

[a] Misdiagnoses made in a large outbreak of botulism.[128]

[b] Paresthesias are reported in approximately 14% of botulism patients (Table 7.1).

toxin type could be determined for 786; of these, 52% were type A, 22% were type B, 25% were type E, and less than 1% were type F.[1] Type A botulism is most common west of the Mississippi River, type B is most common east of the Mississippi, and type E botulism predominates in Alaska.[1,101] This distribution corresponds to the distribution of *C. botulinum* spores in the soil.[102–105] Only three outbreaks of type F botulism have been reported in the U.S. These epidemiologic findings might be helpful retrospectively, distinguishing intentional from foodborne botulism.

LABORATORY FINDINGS

The diagnosis of botulism should be based on the history and physical findings, as routine lab tests are generally nonspecific, and specific confirmation takes days. Routine laboratory tests are not particularly helpful in confirming the clinical suspicion of botulism. The complete blood count, electrolyte panel, renal and liver function tests, urinalysis, and electrocardiogram will all be normal unless complications have occurred. The cerebrospinal fluid (CSF) is typically normal in botulism, whereas the CSF protein is usually elevated in Guillain Barré. The Tensilon test is usually, but not always (see ref. 128) normal in botulism, and can be helpful in distinguishing botulism from myasthenia gravis. The computed tomographic (CT) scan of the head is also normal in botulism, and can be used to rule out stroke, or other intracranial diseases.

Patients with botulism have normal motor nerve conduction velocities and distal latencies. The electromyogram (EMG), however, may be helpful in the diagnosis of botulism and in distinguishing it from other neuromuscular diseases such as myasthenia gravis and Guillain Barré.[106–109] The EMG of involved muscle groups reveals decreased amplitude of the muscle action potential and facilitation during rapid repetitive or post-tetanic stimulation, as can also be seen in patients with Eaton-Lambert syndrome.

Specific laboratory confirmation requires the demonstration of toxin in the blood or gastrointestinal (GI) tract and culturing of *Clostridial* species from the GI tract.[110] Currently, such testing is available only at the CDC and approximately 20 state and municipal public health laboratories.[1] The most sensitive assay for neurotoxin is the mouse bioassay, which is performed by injecting mice intraperitoneally with the toxin containing sample (serum, stool, food extract, etc.) plus or minus polyclonal and type specific antitoxin. The mice are observed for 4 days for the development of botulism, with the mice usually dying from botulism within 6 to 96 hours. Protection by simultaneous administration of antitoxin enables determination of serotype. The mouse bioassay can detect as little as 33 pg of toxin,[5] the mouse LD50. *In vitro* tests to detect botulinum neurotoxin, especially variants of enzyme-linked immunosorbent assays (ELISA) are under development but have not been validated[111–113] and have sensitivities less than that of the mouse bioassay. Simultaneously with collection of blood for serologic studies, stool or wound fluid can be cultured for the presence of *Clostridial* species in the cases of botulism occurring via the oral route.

Unfortunately, current laboratory tests are not particularly sensitive for the detection of botulism. Clostridial cultures were positive for 51% of stool specimens collected from 309 patients with clinically suspected botulism.[89] Toxin testing was positive in only 37% of sera and 23% of stool specimens. At least one laboratory test was positive in 65% of patients.[89] Collecting samples early in the course of disease increases the likelihood of positive results. However large outbreaks have occurred in which no specimens, or a low percentage of specimens, gave positive results.[114] It can also take days for cultures or toxin testing results to be available. It may be possible to culture Clostridial species or detect toxin in source material, if available. Since toxin prepared by terrorists is likely to be crude and unpurified, it is possible that Clostridial nucleic acids may be present on bioterror toxin preparations which could be amplified by PCR for analysis. Cultures and nucleic acid testing would allow more specific classification of the precise Clostridial strain utilized. Since specific therapy with antitoxin must be administered as rapidly as possible to be effective (see below),[115] specific toxin therapy must be based on the clinical diagnosis prior to laboratory confirmation.

TREATMENT

Treatment of botulism includes: (1) early administration of botulinum antitoxin to prevent progression of moderate illness or reduce the duration of mechanical ventilation in patients with rapidly progressive severe botulism, (2) close monitoring of respiratory function (vital capacity and maximal inspiratory force, MIF), and (3) intensive care for patients with significant paralysis and evidence of respiratory insufficiency or failure. Vital capacity (VC) should be measured as soon as the diagnosis is suspected and followed closely. In one study, 10 of 11 patients requiring mechanical ventilation had vital capacities less than 30% of predicted.[116] Patients with VC less than 10 cc/kg should be monitored in an ICU for progression of respiratory failure requiring mechanical ventilation.[117] Patients requiring mechanical ventilation obviously merit ICU care.

Treatment with antitoxin is the mainstay of therapy. In the U.S., more than 80% of adults are treated with antitoxin. Antitoxin is most effective when administered early in the course of disease.[115] Once the toxin has entered the nerve terminal, antitoxin cannot bind.[9] Thus antitoxin only works on circulating toxin. Antitoxin is immunoglobulin harvested from hyperimmunized horses (horse or equine antitoxin) from which the Fc portion has been enzymatically removed (despeciation) to reduce the incidence of side effects such as serum sickness and hypersensitivity reactions. The current licensed equine antitoxin is bivalent, having activity against serotypes A and B. For other toxin serotypes, there is an investigational equine E antitoxin for civilian use and an investigational heptavalent (A–G) antitoxin held by the U.S. Army.[118] The CDC should be contacted for information regarding these products. Efforts are underway to generate human monoclonal antibodies which could replace equine antitoxin therapy.[119–122]

Despite despeciation, there is a significant (9%) risk of hypersensitivity reactions when administering equine antitoxin, including anaphylactic shock.[123] Therefore it is important that skin testing is performed prior to systemic administration according to the protocol provided on the package insert. The amount of antitoxin in one 10-ml vial of antitoxin administered intravenously is enough to neutralize toxin amounts many times in excess of those observed in patients with botulism. The CDC currently recommends a single 10-ml dose of antitoxin, unlike what is suggested in the package insert, due to the 5–8-day reported half-life of the antitoxin.[9] Antitoxin can be obtained via the CDC by contacting the local health department and is diluted 1 : 10 in 0.9% saline and administered slowly by the intravenous route. If the local health department is unavailable, the CDC can be contacted directly at (770) 488–7100. Clinicians should review the package insert with the public health authorities before using antitoxin.

Equine antitoxin has been administered rarely to infants with botulism, because of the risk of lifelong hypersensitivity to equine antigens.[124] In addition, there is some evidence that anaphylaxis may be more severe in infants given equine antitoxin. As an alternative, S. Arnon and colleagues at the Infant Botulism Treatment and Prevention Program at the California Department of Health Services (CDHS) developed human immunoglobulin prepared from volunteers immunized with the investigational botulinum toxoid vaccine. This product, termed botulism immune globulin (BIG-IV) has been evaluated in a prospective randomized trial in infant botulism. Infants with the clinical diagnosis of botulism were randomized to receive either nonimmune human globulin or BIG-IV. Compared to nonimmune globulin, BIG-IV significantly reduced the duration and cost of hospitalization, and the duration of mechanical ventilation and tube feedings.[93] This benefit appeared to accrue even in patients treated as late as 5 to 7 days after the onset of symptoms, an effect due to either ongoing toxin production, and/or slow clearance of toxin from the blood (S. Arnon, personal communication). FDA approval of BIG-IV has been applied for, and BIG-IV may be available for treatment of infant botulism from the CDHS at (510) 540-2646 under an FDA-approved protocol.

Botulism is a reportable disease, and suspected cases should be reported immediately to the hospital epidemiologist or infection control practitioner, as well as the local and state health departments. The phone number of the health department can usually be found in the phone directory under government listings, or via the internet at http://www.cdc.gov/other.htm#states. If local or state health departments are not reachable, the CDC can be contacted directly at (770) 488-7100. For laboratory workers who might be exposed to large amounts of botulinum toxin, an investigational pentavalent (serotypes A, B, C, D, and E) toxoid can be obtained from the CDC.[125] Further details can be obtained from the National Botulism Surveillance and Reference Laboratory at (404) 639-3867. A recombinant vaccine based on the toxin binding domain for many of the serotypes is also under development.[126]

RICIN

Ricin is a plant carbohydrate-binding protein (lectin) found in high concentration in castor beans. It is also cytotoxic, with a typical human lethal dose of approximately 5 μg/kg.[127] While much less toxic than botulinum toxin or staphylococcal enterotoxin B (SEB), ricin can be prepared in liquid or crystalline form in large quantities from castor beans with minimal sophistication or technical capacity. Access to the starting material is simple, since more than one million tons of castor beans are processed annually to produce castor oil; the waste mash contains 5% ricin by weight. Ricin is active orally or upon inhalation and thus could be used to poison food or aerosolized. Its relatively low potency makes it less likely for contaminating water supplies or large volumes of beverages. Ricin has already achieved some notoriety as a poison, being administered via an umbrella tip for a political assassination.[128] A recent find of ricin and castor bean extraction equipment during a police raid of a flat in the United Kingdom indicates that this bioweapon has drawn the attention of terrorists.[129]

RICIN STRUCTURE AND ACTION

Ricin consists of two polypeptide chains connected by a single disulfide bond.[130] The B chain of ricin folds into two globular domains,[131] each of which contains a carbohydrate binding site for β-D-galactopyranoside moieties. These carbohydrates can be present in millions of copies per cell. After binding of ricin to carbohydrate receptors on the cell surface, the toxin is endocytosed and then translocated to the cytosol. In the cytosol, the A chain inactivates ribosomes by catalyzing the removal of an adenine residue from a loop of the 28S ribosome, leading to rapid RNA hydrolysis.[132,133] This results in failure of protein synthesis.

CLINICAL SIGNS AND SYMPTOMS

The clinical findings of ricin intoxication vary depending on the route of exposure, and are largely based on animal studies.[134] After inhalation exposure, there is an incubation period of approximately 8 hours in experimental animals and 4 to 8 hours after an accidental human exposure. In the human exposure, findings included fever, chest tightness, cough, dyspnea, nausea, and arthralgias. In animals, and presumably in humans, larger doses lead to increased permeability pulmonary edema, respiratory failure, shock, and death within 36–72 hours. Inhalational ricin poisoning could mimic a large range of diseases causing acute pulmonary disease including pneumonias and the adult respiratory distress syndrome (ARDS). By the oral route, necrosis of the GI tract will occur, with gastrointestinal, splenic, hepatic, and renal bleeding. The key to diagnosis, as with most of the biothreat agents, is a high index of suspicion and recognition of a cluster of clinically related cases. Routine laboratory tests will show only nonspecific findings. A specific serum ELISA has been described, and acute and convalescent sera can be collected for detection of ricin-specific antibodies. Therapy is supportive, frequently requiring hospitalization in an ICU. No specific approved antitoxin or vaccine exists. Experimental vaccines are under development.[135,136]

STAPHYLOCOCCAL ENTEROTOXIN B

Staphylococcal enterotoxins are a superfamily of proteins secreted by *Stapylococcus aureus* consisting of at least 11 members.[137] Members of the family cause toxic shock syndrome and scalded skin syndrome, while staphylococcal enterotoxin is the most common cause of food poisoning in the U.S. While SEB ingested orally can cause severe gastrointestinal symptoms, the low fatality rate from food poisoning suggests that aerosol release would be a greater biothreat than food or water contamination.

SEB STRUCTURE AND ACTION

The structure of SEB has been determined at 1.5 A resolution.[138] The protein has two binding sites, a high-affinity site for major histocompatibility II (MHC

II) molecules, and a low-affinity site for the Vβ T-cell receptor. As a result, SEB is capable of crosslinking up to 20% of T cells, leading to massive activation of a pro-inflammatory response and release of pro-inflammatory cytokines. The symptoms that result reflect this broad activation of the immune system. As little as 250–400 nanograms of SEB can induce symptoms.[139]

CLINICAL SIGNS AND SYMPTOMS

There is a 1–4-hour incubation period after oral ingestion.[139,140] Usual symptoms include nausea, vomiting, abdominal cramping, and diarrhea. Symptoms typically last up to 20 hours. With severe intoxication, there can be profound dehydration due to loss of fluids, shock, respiratory failure, and cardiovascular collapse. Approximately 15% of patients require hospitalization, with a 5% fatality rate, usually in the very young or the very old. Inhalational SEB exposure would present with different symptoms than seen with oral ingestion (food poisoning). Cytokine activation in the lungs would lead to low-pressure pulmonary edema and acute respiratory failure. The key to diagnosis, as with other biothreat agents, is a high index of suspicion and recognition of a cluster of clinically related cases. Intentional oral release of SEB would have to be distinguished from a naturally occurring outbreak of food poisoning. Routine laboratory tests will show only nonspecific findings. A specific serum ELISA has been described which could be performed on food or environmental specimens or serum, and acute and convalescent sera can be collected for detection of SEB-specific antibodies. Therapy is supportive, but only infrequently requires hospitalization. No specific approved antitoxin or vaccine exists; however, experimental vaccines are under development.[141,142]

CLOSTRIDIA PERFRINGENS EPSILON TOXIN

C. perfringens epsilon toxin is a potential biothreat agent; no reported cases of human disease have occurred. The toxin is produced by *C. perfringens* as a 311-amino-acid protoxin which is cleaved into a 14-amino-acid peptide which is a potent necrotizing toxin. The toxin causes a rapidly fatal toxemia in herbivores when their gastrointestinal tracts are colonized by *C. perfringens*, leading to *in situ* toxin production.[143] The toxin causes pulmonary edema, renal failure, and cardiovascular collapse. The lethal dose for rodents is 100 ng/kg, and it has been estimated that a lethal human dose would be 7 micrograms parenterally.[144] Due to the toxin's potency and lethality, it has been classified as a Category B agent. It is thought that biothreat use would be via the aerosol/inhalational route, leading to pulmonary edema followed by renal failure and cardiovascular collapse.

SUMMARY

In summary, this chapter has described the four major biothreat toxins, botulinum neurotoxin ricin, staphylococcal enterotoxin B, and *Clostridia perfringens* epsilon toxin. The common forensic threat for these threat agents is that the first indication of a bioterror event is likely to be the appearance of intoxicated patients at local hospitals. Rapid appreciation that a biothreat event has occurred is dependent on astute clinicians and a reporting system that can recognize the clustering of cases with a common clinical presentation consistent with intoxication. It is important to recognize that the toxins are all proteins composed of amino acid building blocks. As such, they have a number of forensic features that distinguish them from viral or bacterial threat agents. First, they are not contagious, as the threat agent is not a living organism. For the same reason, these agents cannot be routinely cultured from either patients or the environment after exposure, making forensic detection more difficult. Since proteins are composed of amino acids and not nucleic acid, it is also not possible to amplify and detect the presence of toxins using PCR, or by any type of classic DNA hybridization technology. Rather, detection typically relies on the use of antibodies and serologic testing. Environmental and patient sampling followed by serologic testing is likely to provide the greatest amount of forensic information.

REFERENCES

1. Centers for Disease Control. Botulism in the United States, 1899–1998. *Handbook for epidemiologists, clinicians, and laboratory workers*. Atlanta, Georgia 1998; U.S. Department of Health and Human Services, Public Health Service: downloadable at http://www.bt.cdc.gov/agent/botulism/index,asp.
2. Gill MD. Bacterial toxins: A table of lethal amounts. *Microbiol Rev* 1982;46:86–94.
3. Herrero BA, Ecklung AE, Streett CS, et al. Experimental botulism in monkeys—A clinical pathological study. *Exp Mol Pathol* 1967;6:84–95.
4. Franz DR, Pitt LM, Clayton MA, et al. Efficacy of prophylactic and therapeutic administration of antitoxin for inhalation botulism. *In* DasGupta BR, ed. *Botulinum and Tetanus Neurotoxins: Neurotransmission and Biomedical Aspects*. New York: Plenum Press, 1993. pp. 473–476.
5. Schantz EJ, Johnson EA. Properties and use of botulinum toxin and other microbial neurotoxins in medicine. *Microbiol Rev* 1992;56:80–99.
6. Scott AB, Suzuki D. Systemic toxicity of botulinum toxin by intramuscular injection in the monkey. *Mov Disord* 1988;3:333–335.

7. Arnon SA, Schecter R, Inglesby TV, et al. Botulinum toxin as a biological weapon. *JAMA* 2001;285:1059–1070.
8. Mahant N, Clouston PD, Lorentz IT. The current use of botulinum toxin. *J Clin Neurosci* 2000;7:389–394.
9. Sugiyama H. Clostridium botulinum neurotoxin. *Microbiol Rev* 1980;44(3):419–448.
10. Lacy DB, Stevens RC. Sequence homology and structural analysis of the Clostridial neurotoxins. *J Mol Biol* 1999;291:1091–1104.
11. Middlebrook JL, Franz DR. Botulinum toxin. *In* FR Sidell, ET Takafuji, and DR Franz, DC, eds. *Medical Aspects of Chemical and Biologic Warfare*. Office of the Surgeon General, pp 643–654, 1997.
12. Park JB, Simpson LL. Inhalational poisoning by botulinum toxin and inhalation vaccination with its heavy-chain component. *Infect Immun* 2003;7:1147–1154.
13. Meyer KF. The rise and fall of botulism. *Calif Med* 1973;118:63–64.
14. Midura TF, Arnon SS. Infant botulism. Identification of Clostridium botulinum and its toxins in faeces. *Lancet* 1976;2:934–936.
15. Pickett J, Berg B, Chaplin E, Brunstetter-Shafer MA. Syndrome of botulism in infancy: Clinical and electrophysiologic study. *N Engl J Med* 1976;295:770–772.
16. Davis JB, Mattman LH, Wiley M. Clostridium botulinum in a fatal wound infection. *JAMA* 1951;146:934–936.
17. Werner SB, Passaro D, McGee J, et al. Wound botulism in California, 1951–1998: Recent epidemic in heroin injectors. *Clin Infect Dis* 2000;3:1018–1024.
18. Centers for Disease Control and Prevention, *Botulism-United States, 1978. MMWR*, 1978; 28:73–75.
19. Chia JK, Clark JB, Ryan CA, Pollack M. Botulism in an adult associated with food-borne intestinal infection with Clostridium botulinum. *N Engl J Med* 1986;315:239–241.
20. Griffin PM, Hatheway CL, Rosenbaum RB, Sokolow R. Endogenous antibody production to botulinum toxin in an adult with intestinal colonization botulism and underlying Crohn's disease. *J Infect Dis* 1997;175:633–637.
21. Holzer E. Botulismus durch inhalation. *Med Klin* 1962;41:1735–1740.
22. Cobb DB, Watson WA, Fernandez MC. Botulism-like syndrome after injection of botulinum toxin. Vet. Hum. *Toxicol* 2000;42:163.
23. United Nations Security Council. Tenth report of the Executive Committee of the Special Commission established by the Secretary-General pursuant to paragraph 9(b)(I) of Security Council Resolution 687 (1991), and paragraph 3 of resolution 699 (1991) on the activities of the Special Commission. United Nations Security Council, 1995.
24. Bozheyeva G, Kunakbayev Y, Yeleukenov D. Former Soviet biological weapons facilities in Kazakhstan: Past, present, and future. Center for Nonproliferation Studies, Monterey Institute of International Studies 1999, Monterey, CA.
25. Zilinskas RA. Iraq's biological weapons: The past as future? *JAMA* 1997;278:418–424.
26. St. John R, Finlay B, Blair C. Bioterrorism in Canada: An economic assessment of prevention and postattack exposure. *Can J Infect Dis* 2001;12:275–284.
27. Hatheway CL. Toxigenic Clostridia. *Clin Microbiol Rev* 1990;3:66–98.
28. Hatheway CL. Botulism: The present status of the disease. *Curr Top Microbiol Immunol* 1995;195:55–75.
29. Aureli P, Fenicia L, Pasolini B, et al. Two cases of type E infant botulism caused by neurotoxigenic *Clostridium butyricum* in Italy. *J Infect Dis* 1986;154:207–211.
30. McCroskey LM, Hatheway CL, Fenicia L, et al. Characterization of an organism that produces type E botulinal toxin but which resembles Clostridium butyricum from the feces of an infant with type E botulism. *J Clin Microbiol* 1986;23:201–202.

31. Hall JD, McCroskey LM, Pincomb BJ, Hatheway CL. Isolation of an organism resembling *Clostridium barati* which produces type F botulinal toxin from an infant with botulism. *J Clin Microbiol* 1985;21:654–655.
32. McCroskey LM, Hatheway CL, Woodruff BA, et al. Type F botulism due to neurotoxigenic Clostridium baratii from an unknown source in an adult. *J Clin Microbiol* 1991;29: 2618–2620.
33. Borland ED, Moryson CJ, Smith GR. Avian botulism and the high prevalence of *Clostridium botulinum* in the Norfolk Broads. *Vet Rec* 1977;100:106–109.
34. Galvin JW, Hollier TJ, Bodinnar KD, Bunn CM. An outbreak of botulism in wild waterbirds in southern Australia. *J Wildl Dis* 1985;21:347–350.
35. Jensen WI, Price JI. The global importance of type C botulism in wild birds. *In* Eklund MW, ed. *Avian Botulism: An International Perspective*. Springfield, Ill.: Charles C. Thomas, 1987. pp. 33–54.
36. Smith GR, Milligan RA. Clostridium botulinum type D in Britain. *Vet Rec* 1977;100:121–122.
37. Demarchi J, Mourgues C, Orio J, Prevot AR. Existence du botulisme de type D. *Bull Acad Nat Med* 1958;142:580–582.
38. Prevot AR, Terrasse J, Daumail J, et al. Existence en France du botulisme humain de type C. *Bull Acad Med* (Paris) 1955;139:355–358.
39. Gimenez DF, Ciccarelli AS. Studies on strain 84 of Clostridium botulinum. *Zentralbl Bakteriol* [Orig] 1970;215:212–220.
40. Gimenez DF, Ciccarelli AS. Clostridium botulinum in Argentina: Present and future. *Rev Asoc Argent Microbiol* 1976;8:82–91.
41. Sonnabend O, Sonnabend W, Heinzle R, et al. Isolation of *Clostridium botulinum* type G and identification of type G botulinal toxin in humans: report of five sudden unexpected deaths. *J Infect Dis* 1981;143:22–27.
42. Suen JC, Hatheway CL, Steigerwalt AG, Brenner DJ. *Clostridium argentinense*, sp. nov: A genetically homogeneous group composed of all strains of *Clostridium botulinum* toxin type G and some nontoxigenic strains previously identified as *Clostridium subterminale* or *Clostridium hastiforme*. *Int J Syst Bacteriol* 1988;38:375–381.
43. Hatheway CL, McCroskey LM, Lombard GL, Dowell VR, Jr. Atypical toxin variant of Clostridium botulinum type B associated with infant botulism. *J Clin Microbiol* 1981; 14:607–611.
44. Hatheway CL, McCroskey LM. Examination of feces and serum for diagnosis of infant botulism in 336 patients. *J Clin Microbiol* 1987;25:2334–2338.
45. Gimenez DF. Clostridium botulinum Ba. *Zentralbl Bakteriol Hyg* 1984;257:68–72.
46. Santos-Buelga JA, Collins MD, East AK. Characterization of the genes encoding the botulinum neurotoxin complex in a strain of Clostridium botulinum producing type B and F neurotoxins. *Curr Microbiol* 1998;37:312–318.
47. Dasgupta BR, Boroff DA, Rothstein E. Chromatographic fractionation of the crystalline toxin of Clostridium botulinum type A. *Biochem Biophys Res Commun* 1966;22:750–756.
48. DasGupta BR, Boroff DA. Separation of toxin and hemagglutinin from crystalline toxin of Clostridium botulinum type A by anion exchange chromatography and determination of their dimensions by gel filtration. *J Biol Chem* 1968;243:1065–1072.
49. Balding P, Gold ER, Boroff DA, Roberts TA. Observations on receptor specific proteins. II. Haemagglutination and haemagglutination-inhibition reactions of Clostridium botulinum types A, C, D and E haemagglutinins. *Immunology* 1973;25:773–782.
50. East AK, Stacey JM, Collins MD. Cloning and sequencing of a hemagglutinin component of the botulinum neurotoxin complex encoded by the *Clostridium botulinum* types A and B. *Syst Appl Microbiol* 1994;17:306–312.

51. Fujita R, Fujinaga Y, Inoue K, et al. Molecular characterization of two forms of nontoxic-nonhemagglutinin components of Clostridium botulinum type A progenitor toxins. *FEBS Lett* 1995;376:41–44.
52. East AK, Bhandari M, Stacey JM, et al. Organization and phylogenetic interrelationships of genes encoding components of the botulinum toxin complex in proteolytic Clostridium botulinum types A, B, and F: evidence of chimeric sequences in the gene encoding the nontoxic nonhemagglutinin component. *Int J Syst Bacteriol* 1996;46:1105–1112.
53. Ohishi I, Sugii S, Sakaguchi G. Oral toxicities of Clostridium botulinum toxins in response to molecular size. *Infect Immun* 1977;16:107–109.
54. Chen F, Kuziemko GM, Stevens RC. Biophysical characterization of the stability of the 150-kilodalton botulinum toxin, the nontoxic component, and the 900-kilodalton botulinum toxin complex species. *Infect Immun* 1998;66:2420–2425.
55. Willems A, East AK, Lawson PA, Collins MD. Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A associated with infant botulism: comparison with other Clostridial neurotoxins. *Res Microbiol* 1993;144:547–556.
56. Thompson DE, Brehm JK, Oultram JD, et al. The complete amino acid sequence of the Clostridium botulinum type A neurotoxin, deduced by nucleotide sequence analysis of the encoding gene. *Eur J Biochem* 1990;189:73–81.
57. Binz T, Kurazono H, Wille M, et al. The complete sequence of botulinum neurotoxin type A and comparison with other Clostridial neurotoxins. *J Biol Chem* 1990;265:9153–9158.
58. Whelan SM, Elmore MJ, Bodsworth NJ, et al. Molecular cloning of the Clostridium botulinum structural gene encoding the type B neurotoxin and determination of its entire nucleotide sequence. *Appl Environ Microbiol* 1992;58:2345–2354.
59. Ihara H, Kohda T, Morimoto F, et al. Sequence of the gene for Clostridium botulinum type B neurotoxin associated with infant botulism, expression of the C-terminal half of heavy chain and its binding activity. *Biochim Biophys Acta* 2003;1625:19–26.
60. Hauser D, Eklund MW, Kurazono H, et al. Nucleotide sequence of Clostridium botulinum C1 neurotoxin. *Nucleic Acids Res* 1990;18:4924.
61. Binz T, Kurazono H, Popoff MR, et al. Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type D. *Nucleic Acids Res* 1990;18:5556.
62. Whelan SM, Elmore MJ, Bodsworth NJ, et al. The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin, derived by nucleotide-sequence analysis of the encoding gene. *Eur J Biochem* 1992;204:657–667.
63. Thompson DE, Hutson RA, East AK, et al. Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin: Comparison with other Clostridial neurotoxins. *FEMS Microbiol Lett* 1993;108:175–182.
64. Campbell K, Collins MD, East AK. Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium argentinense) type G neurotoxin: Genealogical comparison with other Clostridial neurotoxins. *Biochim Biophys Acta* 1993;1216:487–491.
65. Gibson AM, Modi NK, Roberts TA, et al. Evaluation of a monoclonal antibody-based immunoassay for detecting type A Clostridium botulinum toxin produced in pure culture and an inoculated model cured meat system. *J Appl Bacteriol* 1987;63:217–226.
66. Kozaki S, Kamata Y, Nishiki T, et al. Characterization of Clostridium botulinum type B neurotoxin associated with infant botulism in Japan. *Infect Immun* 1998;66:4811–4816.
67. Lacy DB, Tepp W, Cohen AC, et al. Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nature Struct Biol* 1998;5:898–902.
68. Hanson MA, Stevens RC. Cocrystal structure of synaptobrevin-II bound to botulinum neurotoxin type B at 2.0 A resolution. *Nat Struct Biol* 2000;7:687–692.

69. Eswaramoorthy S, Kumaran D, Swaminathan S. Crystallographic evidence for doxorubicin binding to the receptor-binding site in Clostridium botulinum neurotoxin B. *Acta Crystallogr D Biol Crystallogr* 2001;57:1743–1746.
70. Dolly JO, Black J, Williams RS, Melling J. Acceptors for botulinum neurotoxin reside on motor nerve terminals and mediate its internalization. *Nature* 1984;307:457–460.
71. Montecucco C. How do tetanus and botulinum toxins bind to neuronal membranes? *Trends Biochem Sci* 1986;11:315–317.
72. Simpson LL. Kinetic studies on the interaction between botulinum toxin type A and the cholinergic neuromuscular junction. *J Pharmacol Expt Ther* 1980;212:16–21.
73. Nishiki T, Tokuyama Y, Kamata Y, et al. The high-affinity binding of Clostridium botulinum type B neurotoxin to synaptotagmin II associated with gangliosides GT1b/GD1a. *FEBS Lett* 1996;378:253–257.
74. Schiavo G, Benfenati F, Poulain B, et al. Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* 1992;359:832–835.
75. Schiavo G, Rossetto O, Catsicas S, et al. Identification of the nerve terminal targets of botulinum neurotoxin serotypes A, D, and E. *J Biol Chem* 1993;268:23784–23787.
76. Schiavo G, Santucci A, Dasgupta BR, et al. Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct COOH-terminal peptide bonds. *FEBS Lett* 1993;335:99–103.
77. Blasi J, Chapman ER, Link E, et al. Botulinum neurotoxin A selectively cleaves the synaptic protein SNAP-25. *Nature* 1993;365:160–163.
78. Binz T, Blasi J, Yamasaki S, et al. Proteolysis of SNAP-25 by types E and A botulinal neurotoxins. *J Biol Chem* 1994;269:1617–1620.
79. Schiavo G, Shone CC, Rossetto O, et al. Botulinum neurotoxin serotype F is a zinc endopeptidase specific for VAMP/synaptobrevin. *J Biol Chem* 1993;268:11516–11519.
80. Schiavo G, Malizio C, Trimble WS, et al. Botulinum G neurotoxin cleaves VAMP/synaptobrevin at a single Ala-Ala peptide bond. *J Biol Chem* 1994;269:20213–20216.
81. Yamasaki S, Hu Y, Binz T, et al. Synaptobrevin/vesicle-associated membrane protein (VAMP) of Aplysia californica: Structure and proteolysis by tetanus toxin and botulinal neurotoxins type D and F. *Proc Natl Acad Sci* 1994;9:4688–4692.
82. Yamasaki S, Binz T, Hayashi T, et al. Botulinum neurotoxin type G proteolyses the Ala81-Ala82 bond of rat synaptobrevin 2. *Biochem Biophys Res Commun* 1994;200:829–835.
83. Blasi J, Chapman ER, Yamasaki S, et al. Botulinum neurotoxin C1 blocks neurotransmitter release by means of cleaving HPC-1/syntaxin. *Embo J* 1993;12:4821–4828.
84. Foran P, Lawrence GW, Shone CC, et al. Botulinum neurotoxin C1 cleaves both syntaxin and SNAP-25 in intact and permeabilized chromaffin cells: Correlation with its blockade of catecholamine release. *Biochemistry* 1996;35:2630–2636.
85. Sutton RB, Fasshauer D, Jahn R, Brunger AT. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. *Nature* 1998;395:347–353.
86. Merson MH, Dowell VR, Jr. Epidemiologic, clinical and laboratory aspects of wound botulism. *N Engl J Med* 1973;289:1005–1010.
87. Wilson R, Morris JG, Jr., Snyder JD, Feldman RA. Clinical characteristics of infant botulism in the United States: a study of the non-California cases. *Pediatr Infect Dis* 1982;1:148–150.
88. Hughes JM, Blumenthal JR, Merson MH, et al. Clinical features of types A and B food-borne botulism. *Ann Intern Med* 1981;95:442–445.
89. Woodruff BA, Griffin PM, McCroskey LM, et al. Clinical and laboratory comparison of botulism from toxin types A, B, and E in the United States, 1975–1988. *J Infect Dis* 1992;166:1281–1286.
90. Maroon JC. Late effects of botulinum intoxication. *JAMA* 1977;238:129.
91. Mann JM, Martin S, Hoffman R, Marrazzo S. Patient recovery from type A botulism: Morbidity assessment following a large outbreak. *Am J Public Health* 1981;71:266–269.

92. Ehrenreich H, Garner CG, Witt TN. Complete bilateral internal ophthalmoplegia as sole clinical sign of botulism: Confirmation of diagnosis by single fibre electromyography. *J Neurol* 1989;236:243–245.
93. Arnon SS. Clinical trial of human botulism immune globulin. *In* DasGupta BR, ed. *Botulinum and Tetanus Neurotoxins: Neurotransmission and Biomedical Aspects*. New York: Plenum Press, 1993. 477–482.
94. de Paiva A, Meunier FA, Molgo J, et al. Functional repair of motor endplates after botulinum neurotoxin type A poisoning: Biphasic switch of synaptic activity between nerve sprouts and their parent terminals. *Proc Natl Acad Sci* U S A 1999;96:3200–3205.
95. Foran PG, Mohammed N, Lisk GO, et al. Evaluation of the therapeutic usefulness of botulinum neurotoxin B, C1, E, and F compared with the long lasting type A. Basis for distinct durations of inhibition of exocytosis in central neurons. *J Biol Chem* 2003;278: 1363–1371.
96. St Louis ME, Peck SH, Bowering D, et al. Botulism from chopped garlic: Delayed recognition of a major outbreak. *Ann Intern Med* 1988;108:363–368.
97. Blake PA, Horwitz MA, Hopkins L, et al. Type A botulism from commercially canned beef stew. *South Med J* 1977;70:5–7.
98. O'Mahony M, Mitchell E, Gilbert RJ, et al. An outbreak of foodborne botulism associated with contaminated hazelnut yoghurt. *Epidemiol Infect* 1990;104:389–395.
99. Seals JE, Snyder JD, Edell TA, et al. Restaurant-associated type A botulism: Transmission by potato salad. *Am J Epidemiol* 1981;113:436–444.
100. Dodds KL. Restaurant-associated botulism outbreaks in North America. *Food Control* 1990; 1:139–141.
101. Eisenberg MS, Bender TR. Botulism in Alaska, 1947 through 1974. Early detection of cases and investigation of outbreaks as a means of reducing mortality. *JAMA* 1976;235:35–38.
102. Meyer KF, Dubovsky BJ. The distribution of the spores of C. botulinum in the United States. *J Infect Dis* 1922;31:559–594.
103. Smith LDS. The occurrence of Clostridium botulinum and Clostridium tetani in the soil of the United States. *Health Lab Sci* 1978;15:74–80.
104. Bott TL, Deffner JS, McCoy E, Foster EM. Clostridium botulinum type E in fish from the Great Lakes. *J Bacteriol* 1966;91:919–924.
105. Ward BQ, Carroll BJ, Garrett ES, Reese GB. Survey of the U.S. Gulf Coast for the presence of Clostridium botulinum. *Appl Microbiol* 1967;15:629–636.
106. Cherington M, Ginsberg S. Type B botulism: Neurophysiologic studies. *Neurology* 1971; 21:43–46.
107. Cherington M. Botulism. Ten-year experience. *Arch Neurol* 1974;30:432–437.
108. Cherington M. Clinical spectrum of botulism. *Muscle Nerve* 1998;21:701–710.
109. Maselli RA, Bakshi N. AAEM case report 16. Botulism. American Association of Electrodiagnostic Medicine. *Muscle Nerve* 2000;23:1137–1144.
110. Dowell VR, Jr., McCroskey LM, Hatheway CL, et al. Coproexamination for botulinal toxin and Clostridium botulinum. A new procedure for laboratory diagnosis of botulism. *JAMA* 1977;238:1829–1832.
111. Emanuel PA, Dang J, Gebhardt JS, et al. Recombinant antibodies: A new reagent for biological agent detection. *Biosens Bioelectron* 2000;14:751–759.
112. Peruski AH, Johnson LH, 3rd, Peruski LF, Jr. Rapid and sensitive detection of biological warfare agents using time-resolved fluorescence assays. *J Immunol Methods* 2002;263:35–41.
113. Ferreira JL, Eliasberg SJ, Harrison MA, Edmonds P. Detection of preformed type A botulinal toxin in hash brown potatoes by using the mouse bioassay and a modified ELISA test. *J AOAC Int* 2001;84:1460–1464.

114. Terranova W, Breman JG, Locey RP, Speck S. Botulism type B: Epidemiologic aspects of an extensive outbreak. *Am J Epidemiol* 1978;108:150–156.
115. Tacket CO, Shandera WX, Mann JM, et al. Equine antitoxin use and other factors that predict outcome in type A foodborne botulism. *Am J Med* 1984;76:794–798.
116. Schmidt-Nowara WW, Samet JM, Rosario PA. Early and late pulmonary complications of botulism. *Arch Intern Med* 1983;143:451–456.
117. Hughes JM, Tacket CO. "Sausage poisoning" revisited. *Arch Intern Med* 1983;143:425–427.
118. Hibbs RG, Weber JT, Corwin A, et al. Experience with the use of an investigational F(ab')2 heptavalent botulism immune globulin of equine origin during an outbreak of type E origin in Egypt. *Clin Infect Dis* 1996;23:337–340.
119. Amersdorfer P, Wong C, Chen S, et al. Molecular characterization of murine humoral immune response to botulinum neurotoxin type A binding domain as assessed using phage antibody libraries. *Infect Immun* 1997;65:3743–3752.
120. Amersdorfer P, Marks JD. Phage libraries for generation of anti-botulinum scFv antibodies. *Methods Mol Biol* 2000;145:219–240.
121. Amersdorfer P, Wong C, Smith T, et al. Genetic and immunological comparison of anti-botulinum type A antibodies from immune and non-immune human phage libraries. *Vaccine* 2002;20:1640–1648.
122. Nowakowski A, Wang C, Powers DB, et al. Potent neutralization of botulinum neurotoxin by recombinant oligoclonal antibody. *Proc Natl Acad Sci* 2002;99:11346–11350.
123. Black RE, Gunn RA. Hypersensitivity reactions associated with botulinal antitoxin. *Am J Med* 1980;69:567–570.
124. Arnon SS. Infant botulism. *In* Feigen R, Cherry J, eds. *Textbook of Pediatric Infectious Diseases*. Philadelphia: W.B. Saunders, 1992. 1095–1102.
125. Siegel LS. Human immune response to botulinum pentavalent (ABCDE) toxoid determined by a neutralization test and by an enzyme-linked immunosorbent assay. *J Clin Microbiol* 1988;26:2351–2356.
126. Byrne MP, Smith LA. Development of vaccines for prevention of botulism. *Biochimie* 2000;82:955–966.
127. Rosenbloom M, Leikin JB, Vogel SN, Chaudry ZA. Biological and chemical agents: A brief synopsis. *Am J Ther* 2002;9:5–14.
128. Knight B. Ricin—A potent homicidal poison. *Br Med J* 1979;1:350–351.
129. Mayor S. UK doctors warned after ricin poison found in police raid. *Br Med J* 2003;326:126.
130. Olsnes S, Kozlov JV. Ricin. *Toxicon* 2001;39:1723–1728.
131. Rutenber E, Robertus JD. Structure of ricin B-chain at 2.5 A resolution. *Proteins* 1991; 10:260–269.
132. Endo Y, Mitsui K, Motizuki M, Tsurugi K. The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and the characteristics of the modification in 28 S ribosomal RNA caused by the toxins. *J Biol Chem* 1987;262:5908–5912.
133. Endo Y, Tsurugi K. RNA N-glycosidase activity of ricin A-chain. Mechanism of action of the toxic lectin ricin in eukaryotic ribosomes. *J Biol Chem* 1987;262:8128–8130.
134. Greenfield RA, Brown BR, Hutchins JB, et al. Microbiological, biological, and chemical weapons of warfare and terrorism. *Am J Med Sci* 2002;323:326–340.
135. Smallshaw JE, Firan A, Fulmer JR, et al. A novel recombinant vaccine which protects mice against ricin intoxication. *Vaccine* 2002;20:3422–3427.
136. Griffiths GD, Phillips GJ, Bailey SC. Comparison of the quality of protection elicited by toxoid and peptide liposomal vaccine formulations against ricin as assessed by markers of inflammation. *Vaccine* 1999;17:2562–2568.
137. Dyer DW, Iandolo JJ. The staphylococcal enterotoxins—A genetic overview. *J Food Safety* 1982;3:249–264.

138. Papageorgiou AC, Tranter HS, Acharya KR. Crystal structure of microbial superantigen staphylococcal enterotoxin B at 1.5 A resolution: implications for superantigen recognition by MHC class 2 molecules and T-cell receptors. *J Mol Biol* 1998;277:61–79.
139. Evenson ML, Hinds MW, Bernstein RS, et al. Estimation of human dose of staphylococcal enterotoxin A from a large outbreak of staphylococcal food poisoning involving chocolate milk. *Int J Food Microbiol* 1988;7:311–316.
140. Holmberg SD, Blake PA. Staphylococcal food poisoning in the United States. New facts and old misconceptions. *JAMA* 1984;251:487–489.
141. LeClaire RD, Hunt RE, Bavari S. Protection against bacterial superantigen staphylococcal enterotoxin b by passive vaccination. *Infect Immun* 2002;70:2278–2281.
142. Boles JW, Pitt LM, LeClaire RD, et al. Generation of protective immunity by inactivated recombinant staphylococcal enterotoxin B vaccine in nonhuman primates and identification of correlates of immunity. *Clin Immun* 2003;108:51–59.
143. McDonel JL. *Toxins of Clostridium perfringens types A, B, C, D, and E.* Oxford: Pergamon Press, 1986.
144. Miyamoto O, Minami J, Toyoshima T, et al. Neurotoxicity of Clostridium perfringens epsilon-toxin for the rat hippocampus via the glutaminergic system. *Infect Immun* 1998;66: 2501–2508.

CHAPTER 8

Epidemiologic Investigation for Public Health, Biodefense, and Forensic Microbiology

STEPHEN A. MORSE

Bioterrorism Preparedness and Response Program, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia

ALI S. KHAN

Division of Parasitic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia

Microorganisms are very efficient at infecting humans, using a number of different strategies and mechanisms. The deliberate dissemination of a biological agent by many of these same mechanisms presents the latest challenge to public health. The deliberate dissemination will often be obvious; however, identifying the covert dissemination of a biological may present challenges. Nonetheless, a thorough investigation integrating epidemiologic data and molecular typing will help to differentiate between a naturally occurring disease outbreak and one resulting from an act of terrorism.

INTRODUCTION

Epidemiology is the study of how disease is distributed in populations and of the factors that influence this distribution.[1] More broadly, it is the study of the distribution and determinants of health-related states or events in specified populations and the application of this study to control health problems.[2] Epidemiology is based on the premise that disease, illness, and ill health are not randomly distributed in a population and that individuals have certain characteristics (e.g., genetic or environmental) that predispose to, or protect against, a variety of different diseases. The specific objectives of epidemiology[1] are to: (1) identify the etiology or cause of a disease and the factors that

Microbial Forensics
Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

increase a person's risk for a disease, (2) determine the extent of disease found in the community, (3) study the natural history and prognosis of disease, (4) evaluate new preventive and therapeutic measures and new modes of health care delivery, and (5) provide a foundation for developing public policy and regulations. This chapter will discuss how epidemiology can be used to identify the source of diseases caused by microorganisms or toxins.

DYNAMICS OF DISEASE TRANSMISSION

Disease has been classically described as the result of an epidemiological triad, where disease results from the interaction of the human host, an infectious agent or toxin, and the environment that promotes the exposure.[1] In some instances, an arthropod vector such as a mosquito or tick is involved. Among the assumptions necessary for this interaction to take place is that there is a susceptible host. The susceptibility of the host is influenced by a variety of factors including genetic, nutritional, and immunological factors. The bacteria, viruses, fungi, and parasites responsible for disease can be transmitted in either a direct or indirect fashion (Table 8.1). Different organisms spread in different ways, and the potential of a given organism to spread and produce outbreaks depends on the characteristics of the organism and the route by which it is transmitted from person to person.

Diseases can be defined as endemic, epidemic, and pandemic. **Endemic** can be defined as either the habitual presence of a disease within a given geographical area, or as the usual occurrence of a given disease within such an area. **Epidemic** can be defined as the occurrence in a community or region of disease, clearly in excess of normal expectancy, and derived from a common source or from a propagated source. **Pandemic** refers to a worldwide epidemic. The usual or expected level of a disease is determined through ongoing **surveillance**.

Microorganisms are very efficient at infecting humans, using a number of different strategies and mechanisms. These are exemplified both by the various

TABLE 8.1 Modes of agent transmission (modified from ref. 1)

- Horizontal
  - Direct transmission
    - Contact (person-to-person)
  - Indirect transmission
    - Common vehicle
    - Single exposure
    - Multiple exposures
    - Continuous exposure
- Vertical (transmission from one generation to another)

strategies devised by the microbe to survive prior to infecting a host such as sporulation or harboring in drought-resistant mosquito eggs, and by the various modes of transmission, e.g., direct contact (including large droplets) or indirect contact with fomites, or by insect vectors, and airborne via small particle droplets.[3] Natural experiments, however, have highlighted the true diversity in the abilities of microorganisms to infect humans and animals: *Salmonella* outbreaks due to contaminated alfalfa sprouts[4] and to ice cream made from milk that was contaminated in a tanker that had previously contained raw eggs,[5] legionellosis associated with grocery misters,[6] Severe Acute Respiratory Syndrome (SARS) and Ebola hemorrhagic fever in healthcare facilities,[7,8] the translocation of Rift Valley fever virus from Africa to the Arabian Peninsula and West Nile virus to the U.S.,[9,10] and pneumonic tularemia on Martha's Vineyard from mowing over a rabbit.[11] These few examples are a semblance of the seemingly endless list of novel ways that agents and their vectors are spread. The ability to exploit newly created biological conditions is the hallmark and challenge of emerging infections.[12]

OUTBREAK INVESTIGATION

The occurrence of a disease at more than an endemic level may stimulate an investigation during which investigators may ask three questions. **Who** was attacked by the disease? The answers to this question will help to identify those characteristics of the human host that are closely related to disease risk. **When** did the disease occur? Some diseases occur with a certain periodicity. This question is also addressed by examining trends of disease incidence over time. **Where** did the cases rise? The answers to the previous questions lead to determining the **how** and **why** of an outbreak. Disease is not randomly distributed in time and place. These questions are central to virtually all outbreak investigations. The investigation of an outbreak may be primarily deductive (i.e., reasoning from premises or propositions proved antecedently) or inductive (i.e., reasoning from particular facts to a general conclusion), or it may be a combination of both. Important considerations in the investigation of acute outbreaks of infectious disease include: (1) determining that an outbreak has in fact occurred, (2) defining the population at risk, (3) determining the method of spread and reservoir, and (4) characterizing the agent. The steps commonly used for investigating an outbreak are shown in Table 8.2.

DELIBERATE INTRODUCTION OF A BIOLOGICAL AGENT

Deliberate dissemination of a biological agent via a number of different routes, including air, water, food, and infected vectors presents the latest challenge to

TABLE 8.2 Commonly used steps in the investigation of an infectious disease outbreak (modified from ref. 1)

Step 1. Define the epidemic.

Define the "numerator" (i.e., cases):

- —Clinical features: is the disease known?
- —What are its serologic or cultural aspects?
- —Are the causes partially understood?

Define the "denominator":

- —What is the population at risk of developing disease?
- —Calculate the attack rates.

Step 2. Examine the distribution of cases by the following:

Time: Look for time-place interactions.

Place

Person: Examine the risk in subgroups of the affected population according to personal characteristics: sex, age, residence, occupation, social group, etc.

Step 3. Look for combinations (interactions) of relevant variables.

Step 4. Develop hypotheses based on the following:

Existing knowledge (if any) of the disease

Analogy to diseases of known etiology

Step 5. Test hypotheses.

Further analyze existing data (case-control studies)

Collect additional data

Step 6. Recommend control measures.

Control of present outbreak

Prevention of future similar outbreaks

---

the global public health. The deliberate nature of such dissemination will often be obvious, as in the case of multiple mailed letters containing highly refined anthrax spores. However, some forms of bioterrorism may be more covert. For example, the deliberate contamination of salad bars in The Dalles, Oregon in 1984 by a religious cult in an effort to test their ability to incapacitate the local population prior to an election sickened more than 750 persons.[13] The outbreak was specifically excluded as bioterrorism during the initial investigation, and only recognized as such following a tip from an informant.[14] Given the natural ability of infectious agents to emerge, the Oregon outbreak serves to highlight the difficulties in determining a characteristic signature for an infectious disease outbreak resulting from deliberate transmission.

The difficulties in identifying a covert dissemination of a biological agent are exemplified by the investigation in The Dalles of a foodborne outbreak with a very unusual pattern and a rare strain of *Salmonella typhimurium*. Although the possibility of intentional contamination was considered early in the investigation, it was specifically excluded for the following reasons: (1)

such an event had never been reported previously, (2) no one claimed responsibility, (3) no disgruntled employee was identified, (4) no motive was apparent, (5) the epidemic curve suggested multiple exposures, which was presumed to be unlikely behavior for a saboteur, (6) law enforcement officials failed to establish a recognizable pattern of unusual behavior, (7) a few employees had onset of illness before the patrons, suggesting a possible source of infection, (8) the outbreak was biologically plausible—even if highly unlikely, and, (9) it is not unusual to not be able to find a source in even highly investigated outbreaks. Although one of the initial reasons to exclude terrorism (i.e., no prior incidents) is no longer applicable, based on similar actions since 1984, determining if an unusual outbreak is biologically plausible will remain a challenge. In this context, it is important to remember that the first case of inhalation anthrax identified in Florida in 2001 was initially thought to be natural. It is clear from the two documented cases of bioterrorism in the U.S.—the 1984 Oregon salmonella outbreak and the 2001 anthrax attack—that a terrorist will not necessarily announce his/her intentions or take credit for such an attack.[13,15] Similarly, divining the motives behind an attack should be abandoned as a public health tool to assess whether an outbreak is natural or deliberate in nature. Fortunately, there are a number of epidemiologic clues that in themselves or in combination may suggest that an outbreak is deliberate. It is essential to make this determination not only from the law enforcement standpoint to prevent future such actions, but to protect the public health. There is a very short "window of opportunity" in which to implement postexposure prophylaxis for many of the agents likely to be used for bioterrorism.[16] Even when postexposure prophylaxis may be unavailable or of limited utility, ascertaining the deliberate nature of an attack can allow for more effective postexposure planning for potential casualties and to improve surveillance for additional events. Therefore, it is critical that all outbreaks be rapidly investigated and assessed for whether they are of deliberate origin.

A set of epidemiologic clues (Table 8.3) has been proposed by the Department of Health and Human Service's Centers for Disease Control and Prevention (CDC) in collaboration with the Federal Bureau of Investigation (FBI).[17] These clues are based on distinctive epidemiology and laboratory criteria of varying specificity to evaluate whether an outbreak may be of deliberate origin. The clues focus on aberrations in the typical characterization of an outbreak by person, place, and time in addition to consideration of the microorganism. Some of the clues, such as a community-acquired case of smallpox, are quite specific for bioterrorism whereas others, such as similar genetic typing of an organism, may simply denote a natural outbreak. A combination of clues, especially those that suggest suspicious point source outbreaks, will increase the probability that the event is likely due to bioterrorism. Although these clues

TABLE 8.3 Epidemiologic clues that may signal a biologic or chemical terrorist attack (modified from ref. 17)

1. Single case of disease caused by an uncommon agent (e.g., glanders, smallpox, viral hemorrhagic fever, inhalation or cutaneous anthrax) without adequate epidemiologic explanation
2. Unusual, atypical, genetically engineered, or antiquated strain of an agent (or antibiotic-resistance pattern)
3. Higher morbidity and mortality in association with a common disease or syndrome or failure of such patients to respond to usual therapy
4. Unusual disease presentation (e.g., inhalation anthrax or pneumonic plague)
5. Disease with an unusual geographic or seasonal distribution (e.g., plague in a nonendemic area, influenza in the summer)
6. Stable endemic disease with an unexplained increase in incidence (e.g., tularemia, plague)
7. Atypical disease transmission through aerosols, food, or water, in a mode suggesting sabotage (i.e., no other possible physical explanation)
8. No illness in persons who are not exposed to common ventilation systems (have separate closed ventilation systems) when illness is seen in persons in close proximity who have a common ventilation system
9. Several unusual or unexplained diseases coexisting in the same patient without any other explanation
10. Unusual illness that affects a large, disparate population (e.g., respiratory disease in a large heterogeneous population may suggest exposure to an inhaled pathogen or chemical agent)
11. Illness that is unusual (or atypical) for a given population or age group (e.g., outbreak of measles-like rash in adults)
12. Unusual pattern of death or illness among animals (which may be unexplained or attributed to an agent of bioterrorism) that precedes or accompanies illness or death in humans
13. Unusual pattern of death or illness in humans that precedes or accompanies illness or death in animals (which may be unexplained or attributed to an agent of bioterrorism)
14. Ill persons who seek treatment at about the same time (point source with compressed epidemic curve)
15. Similar genetic type among agents isolated from temporally or spatially distinct sources
16. Simultaneous clusters of similar illness in noncontiguous areas, domestic or foreign
17. Large numbers of cases of unexplained diseases or deaths

are an important set of criteria to help evaluate outbreaks, no list will replace sound epidemiology to assess an outbreak.

It is important to note that epidemiologic clues can only be assessed in the context of a rapid and thorough epidemiologic investigation. Not surprisingly, surveillance to identify increases in disease incidence is both the first step and the cornerstone of bioterrorism epidemiology. The majority of the clues described in Table 8.3 simply suggest an unusual cluster of cases. They have been reorganized by specificity to trigger increasingly broader investigations

TABLE 8.4 Recommendations for level of public health involvement for investigation of potential biologic or chemical terrorism (modified from ref. 17)

Initial investigation at the local level

a. Higher morbidity and mortality than expected, associated with a common disease or syndrome
b. Disease with an unusual geographic or seasonal distribution
c. Multiple unusual or unexplained disease entities coexisting in the same patient
d. Unusual illness in a population (e.g., renal disease in a large population which may be suggestive of toxic exposure to an agent such as mercury)
e. Ill persons seeking treatment at about the same time
f. Illness in persons suggesting a common exposure (e.g., same office building, meal, sporting event, or social event)

Continued investigation with involvement of the state health department and/or Centers for Disease Control and Prevention

a. At least a single, definitively diagnosed case(s) with one of the following:
  —Uncommon agent or disease
  —Illness due to a genetically altered organism
b. Unusual, atypical, or antiquated strain of agent
c. Disease with unusual geographic, seasonal, or "typical patient" distribution
d. Endemic disease with unexplained increase in incidence
e. No illness in persons not exposed to common ventilation systems
f. Simultaneous clusters of similar illness in non-contiguous areas, domestic or foreign
g. Cluster of patients with similar genetic type among agents isolated from temporally or spatially distinct sources

by state and federal public health officials and to alert law enforcement authorities (Tables 8.4 and 8.5). However, even the most specific of clues may signal a new natural outbreak. For example, the recent community outbreak of individuals with smallpox-like lesions in the Midwest may, on first blush, have indicated the deliberate release of smallpox virus. However, a thorough integrated epidemiologic and laboratory investigation identified the disease as monkeypox, an exotic disease in the U.S., which in itself should suggest bioterrorism.[18] Affected individuals were infected by prairie dogs purchased as pets, which had acquired their infection while co-housed with infected Giant Gambian rats that had recently been imported from Ghana, and not from deliberate dissemination. Similarly, other emerging infectious diseases such as West Nile encephalitis and SARS would appropriately meet the criteria for suspect bioterrorism and require a thorough investigation.

MOLECULAR STRAIN TYPING

The microbiology laboratory has made significant contributions to the epidemiology of infectious diseases. The repeated isolation of a specific

TABLE 8.5 Considerations for notifying law enforcement of a possible biologic or chemical terrorism initial investigation at the local level (modified from ref. 17)

Immediate notification of the FBI when:

- a. Notification is received from individual or group that a terrorist attack has occurred or will occur.
- b. A potential dispersal/delivery device such as munition or sprayer or questionable material is found.

Notification of the FBI as soon as possible after an investigation confirms the following:

- a. Illness due to unexplained aerosol, food, or water transmission.
- b. At least a single, definitively diagnosed case(s) with one of the following:
  - —Uncommon agent or disease occurring in a person with no other explanation
  - —Illness due to a genetically altered organism

Notification of the FBI after an investigation confirms the following (with no plausible natural explanation):

- a. Disease with an unusual geographic, seasonal, or "typical patient" distribution
- b. Unusual, atypical, or antiquated strain of agent
- c. Simultaneous clusters of similar illness in noncontiguous areas, domestic or foreign
- d. Clusters of patients presenting with similar genetic type among agents isolated from temporally or spatially distinct sources

---

microorganism from patients with a given disease or syndrome has helped to prove infectious etiologies. In addition, the isolation and identification of microorganisms from animals, vectors, and environmental sources has been invaluable in identifying reservoirs and verifying modes of transmission. In dealing with an infection, it is often necessary to identify the species of the infecting microorganism in order to prescribe effective therapy. Many of the techniques that have evolved for such purposes are both rapid and accurate but, in general, do not provide the kind of genetic discrimination necessary for addressing epidemiologic questions. The epidemiology of many infectious diseases is becoming more complex. Fortunately, typing methods for bacteria, fungi, protozoa, and viruses have evolved to meet this challenge. Historically, the typing methods that have been used in epidemiologic investigations fall into two broad categories: *phenotypic* methods and *genotypic* methods. Phenotypic methods are those methods that characterize the products of gene expression in order to differentiate strains. For example, the use of biochemical profiles to discriminate between genera and species of bacteria is used as a diagnostic method, but can also be used for biotyping. Other methods, such as phage typing, can be used to discriminate among groups within a bacterial species. Biotyping emerged as a useful tool for epidemiologic investigations in the 1960s and early 1970s, while phage typing of bacteria and serological typing of bacteria and viruses has been used for decades. Today, the majority of these tests are considered inadequate for epidemiologic purposes. First, they

TABLE 8.6 Characteristics of phenotypic typing methods (modified from ref. 23)

Typing System	Proportion of Strains Typeable	Reproducibility	Discriminatory Power	Ease of Interpretation	Ease of Performance
Biotyping	All	Poor	Poor	Moderate	Easy
Antimicrobial susceptibility patterns	All	Good	Poor	Easy	Easy
Serotyping	Most	Good	Fair	Moderate	Moderate
Bacteriophage or pyocin typing	Some	Good	Fair	Difficult	Difficult
MLEE[1]	All	Excellent	Excellent	Moderate	Moderate

[1] MLEE, multilocus enzyme electrophoresis.

do not provide enough unrelated parameters to obtain a good reflection of genotype. For example, serotyping of *Streptococcus pneumoniae* discriminates among only a limited number of groups. In addition, some virus species, such as human cytomegalovirus and measles virus, cannot be divided into different types or subtypes by serology, because significant antigenic differences do not exist. Second, the expression of many genes is affected by spontaneous mutations, environmental conditions, and by developmental programs or reversible phenotypic changes, such as high-frequency phenotypic switching. Because of this, many of the properties measured by phenotypic methods have a tendency to vary, and for the most part they have been replaced by genotypic methods. The one major exception is multilocus enzyme electrophoresis (MLEE),[19,20] which is a robust phenotypic method that performs comparably with many of the most effective DNA-based methods.[21,22] Characteristics of selected phenotypic methods are presented in Table 8.6. These methods have been characterized by: *typeability*, which is the ability of the technique to assign an unambiguous result (i.e., type) to each isolate; *reproducibility*, which is when a method yields the same results upon repeat testing of a bacterial strain; *discriminatory power*, which is the ability of the method to differentiate among epidemiologically unrelated isolates; *ease of interpretation*, which refers to the effort and experience required to obtain useful, reliable typing information using a particular method; and *ease of performance*, which reflects the cost of specialized reagents and equipment, technical complexity of the method, and the effort required to learn and implement the method.

Extremely sensitive and specific molecular techniques have recently been developed to facilitate epidemiologic studies. Our ability to use these

TABLE 8.7 Molecular characterization of genetic diversity at different hierarchical levels (modified from ref. 24)

Function	Purpose	Regions of DNA
Discrimination above level of species	Taxonomy/evolution	Highly conserved coding regions (e.g., rDNA)
Discrimination between species	Taxonomy/diagnosis/ epidemiology	Moderately conserved regions
Discrimination between intraspecific variants/strains	Population genetics	Variable regions
Discrimination between individual isolates/clonal lineages	"Fingerprinting"—tracking transmission of genotypes/ identifying sources of infection and risk factors	Highly variable genetic markers that are not under selection by the host
Genetic markers/linking phenotype and genotype	Identifying phenotypic traits of clinical significance	Genotype linked to phenotype

molecular techniques (genotypic methods) to detect and characterize the genetic variability of infectious agents (bacteria, fungi, protozoa, viruses) is the foundation for the majority of molecular epidemiological studies. The application of appropriate molecular techniques has been an aid in the surveillance of infectious agents and in determining sources of infection. These molecular techniques can be used to study health and disease determinants in animal (including human) as well as plant populations. It requires choosing a molecular method(s) that is capable of discriminating genetic variants at different hierarchical levels, coupled with the selection of a region of nucleic acid, which is appropriate to the questions being asked (Table 8.7).

Genotypic methods are those that are based on an analysis of the genetic structure of an organism. Over the past decade, a number of genotypic methods have been used to fingerprint pathogenic microorganisms (Table 8.8). The methods have been described in detail elsewhere.[23–27] Among these methods, RFLP-PFGE (restriction fragment length polymorphism/pulsed-field gel electrophoresis) and RFLP + probe, and ribotyping have been the most commonly used methods for fingerprinting bacteria.[25,28] RAPD (random amplification of polymorphic DNA) and karyotyping have been used for fingerprinting fungi.[25,29] MLEE (multilocus enzyme electrophoresis), RAPD, and PCR (polymerase chain reaction)-RFLP have been used for fingerprinting parasitic protozoa.[25] Select gene or complete genome characterization, as well as other molecular methods, have been used for viruses.[30]

When should fingerprinting be used? Strain typing data are most effective when they are collected, analyzed, and integrated into the results of an epidemiological investigation. The epidemiologist should consult the laboratory

TABLE 8.8 Examples of genotypic methods used in epidemiologic investigations

Restriction endonuclease-based methods
- A. Restriction fragment length polymorphism (RFLP) without hybridization
  - —Frequent cutter (4–6 bp recognition site) coupled with conventional electrophoresis to separate restriction fragments
  - —Infrequent cutter (generally 6–8 bp recognition site) coupled with pulsed-field gel electrophoresis (PFGE) to separate restriction fragments
- B. RFLP with hybridization
  - —Frequent cutter (4–6 bp recognition site) coupled with conventional electrophoresis to separate restriction fragments followed by Southern transfer to nylon membrane. The power and efficacy of typing method depends on the probe.
    - —16S and 23S rRNA (ribotyping)
    - —Insertion sequences (e.g., IS6110 of *Mycobacterium tuberculosis*)

Amplification-based methods
- A. Random amplification of polymorphic DNA (RAPD) analysis; arbitrarily primed PCR (APPCR)
- B. Amplified fragment length polymorphism (AFLP) method
- C. Repetitive element PCR (REP-PCR) method; variable number tandem repeat (VNTR) fingerprinting

Sequence-based methods
- A. Multilocus sequence typing (MLST)
- B. Electrophoretic karyotyping

when investigating a potential outbreak of an infectious disease. Microbial fingerprinting should supplement, and not replace, a carefully conducted epidemiological investigation. In some cases, typing data can effectively rule out an outbreak and thus avoid the need for an extensive epidemiological investigation. In other cases, these data may reveal the presence of outbreaks caused by more than one strain. Data interpretation is facilitated greatly by an appreciation of the molecular basis of genetic variability of the organism being typed and the technical factors that can affect results. With the exception of whole-genome sequencing, the molecular methods analyze only a small portion of the organisms' genetic complement. Thus, isolates that give identical results are classified as "indistinguishable," not "identical." Theoretically, a more detailed analysis should uncover differences in the isolates that appeared to give identical patterns, but that were epidemiologically unrelated. This is unlikely to occur when a set of epidemiologically linked isolates are analyzed.[23] For this reason, only whole-genome sequencing would provide the unequivocal data required for attribution.

The power of molecular techniques in epidemiological investigations is well exemplified by a few examples. PulseNet, the national molecular subtyping network for foodborne disease surveillance, was established by the CDC and several state health departments in 1996 to facilitate subtyping bacterial foodborne pathogens for epidemiologic purposes. Twenty years ago, most foodborne outbreaks were local problems that typically resulted from improper food-handling practices. Outbreaks were often associated with individual restaurants or social events, and often came to the attention of local public health officials through calls from affected persons. Today, foodborne disease outbreaks commonly involve widely distributed food products that are contaminated before distribution, resulting in cases that are spread over several states or countries. The PulseNet network, which began with 10 laboratories typing a single pathogen (*Escherichia coli* O157:H7), has grown and now includes 46 state and two local public health laboratories and the food safety laboratories of the U.S. Food and Drug Administration (FDA) and the U.S. Department of Agriculture (USDA).[28] Currently, four foodborne pathogens (*E. coli* O157:H7, nontyphoidal *Salmonella* serotypes, *Listeria monocytogenes*, and *Shigella*) are being subtyped by PFGE as part of routine surveillance for foodborne disease. The laboratories follow a standardized protocol using similar equipment so that results are highly reproducible and DNA patterns generated at different laboratories can be compared. Isolates are subtyped on a routine basis, and the data analyzed promptly at the local level. Clusters can often be detected locally that could not have been identified by traditional epidemiologic methods alone. PFGE patterns are shared between participating laboratories electronically, which serves to link apparently unrelated outbreaks and facilitates the identification of a common vehicle.[31] For example, in May 1998, PulseNet facilitated the investigation of two clusters of *E. coli* O157:H7 in the northeastern U.S. PFGE fingerprinting of the *E. coli* O157:H7 isolates by the PulseNet laboratories in that region revealed two simultaneous clusters of *E. coli* O157:H7 infections (32 isolates in four of five states with one PFGE pattern, and 25 isolates in all five states with a second pattern), one of which could be traced to two supermarkets that received ground beef from the same distributor. Without molecular typing, epidemiologists would have found it difficult to identify cases associated with each cluster. On the other hand, the use of PFGE subtyping as part of routine surveillance has benefits beyond outbreak detection. For example, the temporal clustering of unrelated cases is not uncommon, and without molecular typing, valuable public health resources would be wasted investigating pseudo-outbreaks.

Another example of the power of molecular techniques in solving an epidemiologic investigation involves a case of HIV transmission by a healthcare worker. The investigation involved a young woman who had contracted AIDS even though she had no identifiable risk factors. During the investigation, it was revealed that 2 years previously she had several teeth extracted by a dentist who was subsequently confirmed as having AIDS. A retrospective case-control study was conducted of the dentist and his former patients to evaluate the possibility of dentist-to-patient transmission. Patients were questioned to ascertain known risk factors for HIV transmission. Infection control practices in the dental office were also evaluated.[32] Eight HIV-positive persons were identified from among a group of more than 1,000 former patients of the dentist. Five of the eight patients had no risk factors or other documented exposures to HIV. Although all five had undergone invasive procedures, and four of the five shared visit days, no identifiable mechanism of transmission could be established by traditional case-control methodology.[32] However, a comparison of the nucleotide sequences of several regions of the gp120 gene of the HIV strains of the dentist, HIV-positive patients (with and without known risk factors), and 35 HIV-infected community controls established the likelihood of a common source of infection.[33] The genetic distance of viruses from the five patients without known risk factors and the virus from the dentist was 3.4%–4.9%, which is similar to that found previously with HIV viruses from persons with epidemiologically linked infections. In contrast, isolates from patients with known risk factors were more distantly related (>10%) to the HIV virus obtained from the dentist.[33] The average genetic distance of viruses from the five patients and the community controls was approximately 11%, which was virtually identical to the average distance among the 35 HIV viruses from controls. Phylogenetic tree analysis confirmed that the HIV viruses from the dentist and the five patients formed a tightly related cluster.[33]

SUMMARY

With few exceptions, it will require a careful epidemiologic investigation to determine whether an outbreak of infectious disease is due to the intentional release of an agent, or is naturally occurring. A number of molecular techniques have been developed for subtyping bacteria, viruses, fungi, and protozoa, which will facilitate this investigation as well as identify clusters of related microorganisms.

REFERENCES

1. Gordis L. Epidemiology. Philadelphia, PA: WB Saunders Co.; 1996.
2. Last JM. *A Dictionary of Epidemiology*, 2nd ed. New York, NY: Oxford University Press; 1988.
3. Chin J (ed). *Control of Communicable Diseases Manual*, 17th ed. Washington, DC: American Public Health Association; 2000.
4. Centers for Disease Control. Outbreak of *Salmonella* serotype Kottbus infections associated with eating alfalfa sprouts—Arizona, California, Colorado, and New Mexico, February–April, 2001. *MMWR Morb Mortal Wkly Rep* 2001;51:7–9.

5. Hennessy TW, Hedberg CW, Slutsker L, White KE, Besser-Wiek JM, Moen ME, et al. A national outbreak of *Salmonella enteritidis* infections from ice cream. *N Engl J Med* 1996;334: 1281–1286.
6. Mahoney FJ, Hoge CW, Farley TA, Barbaree JM, Breiman RF, Benson RF, et al. Community-wide outbreak of Legionnaires' disease associated with a grocery store mist machine. *J Infect Dis* 1992;165:736–739.
7. Centers for Disease Control and Prevention. Cluster of severe acute respiratory syndrome among protected health-care workers—Toronto, Canada, April 2003. *MMWR Morb Mortal Wkly Rep* 2003;52:433–436.
8. Centers for Disease Control. Outbreak of Ebola viral hemorrhagic fever—Zaire, 1995. *MMWR Morb Mortal Wkly Rep* 1995;44:381–382.
9. Centers for Disease Control and Prevention. Outbreak of Rift Valley fever—Saudi Arabia, August-October, 2000. *MMWR Morb Mortal Wkly Rep* 2000;49:905–908.
10. Nash D, Mostashari F, Fine A, Miller J, O'Leary D, Murray K, et al. The outbreak of West Nile virus infection in the New York City area in 1999. *N Engl J Med* 2001;344:1807–1814.
11. Feldman KA, Enscore RE, Lathrop SL, Matyas BT, McGuill M, Schrieffer ME, et al. An outbreak of pneumonic tularemia on Martha's Vineyard. *N Engl J Med* 2001;345:1601–1606.
12. Institute of Medicine. Emerging infections: Microbial threats to the health in the United States. Washington, DC: *National Academy Press* 1994.
13. Torok TJ, Tauxe RV, Wise RP, Livengood JR, Sokolow R, Mauvais S, et al. A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. *JAMA* 1997;278:389–395.
14. Carus WS. Working paper. Bioterrorism and biocrimes. The illicit use of biological agents since 1900. Washington, DC: National Defense University; 2001.
15. Jernigan JA, Stephens DS, Ashford DA, Omenaca C, Topiel MS, Galbraith M, et al. Bioterrorism-related anthrax: The first 10 cases reported in the United States. *Emerg Infect Dis* 2001;7:933–944.
16. Khan AS, Morse S, Lillibridge S. Public health preparedness for biological terrorism in the USA. *Lancet* 2000;356:1179–1182.
17. Treadwell TA, Koo D, Kuker K, Khan AS. Epidemiologic clues to bioterrorism. *Pub Health Rep* 2003;118:92–98.
18. Centers for Disease Control and Prevention. Update: Multistate outbreak of monkeypox—Illinois, Indiana, Kansas, Missouri, Ohio, and Wisconsin, 2003. *MMWR Morb Mortal Wkly Rep* 2003;52:642–646.
19. Pasteur N, Pasteur G, Bonhomme F, Catalan J, Davidian JB. *J Practical isozyme genetics.* New York, NY: Halsted Press; 1988.
20. Richardson BJ, Baverstock PR, Adams M. Alloenzyme electrophoresis, *A Handbook for Animal Systematics and Population Studies*. Orlando, FL: Academic Press, Inc.; 1986.
21. Pujol C, Joly S, Lockhart SR, Noel S, Tibayrenc M, Soll D. Parity among the randomly amplified polymorphic DNA method, multilocus enzyme electrophoresis, and Southern blot hybridization with the moderately repetitive DNA probe Ca3 for fingerprinting *Candida albicans*. *J Clin Micro*1997;35:2348–2358.
22. Tibayrenc M, Neubauer K, Barnabe C, Guerrini F, Skarecky D, Ayala FJ. Genetic characterization of six parasitic protozoa: Parity between random primer DNA typing and multilocus enzyme electrophoresis. *Proc Natl Acad Sci* USA 1993;90:1335–1339.
23. Tenover FC, Arbeit RD, Goering RV. How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: A review for healthcare epidemiologists. *Infect Contr Hosp Epidemiol* 1997;18:426–439.
24. Thompson RCA, Constantine CC, Morgan UM. Overview and significance of molecular methods: What role for molecular epidemiology? *Parasitology* 1998;117:S161–S175.

25. Soll DR, Lockhart SR, Pujol C. Laboratory procedures for the epidemiological analysis of microorganisms. *In* Murray PR, Baron EJ, Jorgensen JH, Pfaller MA, Yolken RH (eds.). *Manual of Clinical Microbiology*, 8th ed. Washington, DC: ASM Press, 2002:139–161.
26. Pennington TH. Electrophoretic typing. *In* Sussman M (ed.). *Molecular Medical Microbiology.* San Diego, CA: Academic Press, 2002:535–547.
27. Arens M. Methods for subtyping and molecular comparison of human viral genomes. *Clin Microbiol Rev* 1999;12:612–626.
28. Swaminathan B, Barrett TJ, Hunter SB, Tauxe RV, and the CDC PulseNet Task Force. PulseNet: The molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerg Infect Dis* 2001;7:382–389.
29. Soll DR. The ins and outs of DNA fingerprinting of infectious fungi. *Clin Microbiol Rev* 2000;13:332–370.
30. Arens M. Methods for subtyping and molecular comparison of human viral genomes. *Clin Microbiol Rev* 1999;12:612–626.
31. Centers for Disease Control and Prevention. Outbreaks of *Shigella sonnei* infection associated with eating fresh parsley—United States and Canada, July–August, 1998. *MMWR Morb Mortal Wkly Rep* 1999;48:285–289.
32. Ciesielski C, Marianos D, Ou CY, et al. Transmission of human immunodeficiency virus in a dental practice. *Ann Intern Med* 1992;116:798–805.
33. Ou CY, Ciesielski CA, Meyers G, et al. Molecular epidemiology of HIV transmission in a dental practice. *Science* 1992;256:1165–1171.

CHAPTER 9

Molecular Epidemiology and Forensics of RNA Viruses

CONSUELO CARRILLO AND DANIEL L. ROCK
Plum Island Animal Disease Center, Agricultural Research Service, U.S. Department of Agriculture, Greenport, NY

Forensic science is interested in detecting a crime, handling evidence correctly, and using that evidence to identify and prosecute the perpetrators. From the microbiological point of view, the goal of forensic analysis of pathogens will not be merely to detect the presence of the agent but also to identify the pathogen, its source, and its origin in a scientifically based unequivocal way that can be used as evidence in court.[1,2]

Compared to other fields of forensic investigation, microbial forensics is for the most part not helped by other ancilliary evidence such as pollen, minerals, dust, hair, etc. Thus, any evidence relating a pathogen to its source and origin must be found in the "genetic language" of the pathogen and the "memory" of the pathogen population.

Modern molecular techniques allow detailed analysis of the genetic code of organisms, permitting distinction of species, strains, and in some cases even individuals. Natural variation in the nucleotide sequence of the organism's genome (sequence polymorphisms) results in genetic differences between closely related individuals. For forensic purposes, highly polymorphic genomic areas are usually attractive, because of their high power of discrimination. However, since these areas undergo frequent mutation and are subject to an unknown number of natural selective pressures, interpretation of these genetic profiles requires knowledge about genetic variability, which includes mutation frequency, mutation rates, mutation and recombination processes, tolerance and constraints of genes and proteins, and evolution and memory of the microbial population. This information is critical for choosing the most informative genomic areas or a combination of them, for analysis.

Although complete genome sequence analysis of viruses has accelerated in recent years, the amount of data available is still rather limited. In most cases,

Microbial Forensics
Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

restricted areas of the viral genome are extensively characterized for routine epidemiological purposes. There is a pressing need to identify all informative areas of the pathogen genome to improve routine epidemiological analysis and meet the new demands of forensic microbiology.

CHALLENGES POSED BY RNA VIRUSES

The low mutation rate of DNA systems allows the use of techniques including single-nucleotide polymorphism (SNP) and deletion/insertion polymorphism (DIP) assays in forensic analysis.[3–13] In addition, a modification of these assays based on melting temperature profiles of the DNA segments that compose the total DNA sample offers a less expensive genotyping method.[14] Multiplex DNA amplification can be optimized and used to produce up to 30 simultaneous DNA targets, mostly consisting of fragments containing SNPs that can be differentiated by size, and used for diagnostic and forensic purposes.[15] Higher mutability and rapid evolutionary capabilities of RNA viral genomes require that other approaches be taken.

Many important viral pathogens have an RNA genome. RNA viral pathogens are abundant in humans, animals, and plants.[16] Over the last decade, work with many RNA viruses has established that they all share high mutability and rapid evolutionary capabilities. Their genetic plasticity may in part explain why they are frequently "emerging pathogens" in new hosts.[17–24]

Due to lack of proofreading of the RNA polymerases and to the high error rate of reverse transcriptases,[25–27] RNA viruses exhibit high mutation rates that result in genetically heterogeneous populations in continuous competition: these are able to select the best-adapted variants for each particular environment in a quasi-deterministic fashion.[28–32] Features of RNA replication are high mutation rates, short replication times, and high progeny yields. These lead to misincorporation of at least 1 nucleotide every 1,000 to 100,000 nucleotides. Mutation rates per nucleotide site in the range of 10^{-3} to 10^{-5} for a 10-kb genome ensure that each progeny genome includes on average 0.1–10 mutations.[30,32–37] As a result, RNA virus populations are composed of a wide variety of individuals, rarely identical to each other but highly similar genetically, in which most individuals conserve the better fitness-adapted genetic information (master sequence).[30,32,38–41] The advantage of this population structure (*quasispecies*) is that the species may overcome almost any change in the environment with just a switch in the master sequence, giving replicative advantage under new conditions to the most-fit individual among minor variants.[42] The quasispecies concept was proposed by Eigen et al. to describe error-prone replication and self-organization of primitive macromolecules.[40,41,43,44] The concept has proven useful for explaining and understanding RNA genetic elements which are subjected to a continuous dynamic of mutation, competition, and selection.[45–50] Application of the quasispecies concept of population distribution and evolution to RNA viruses has also been very useful for developing an understanding of their biology in nature.

A CASE STUDY: FOOT-AND-MOUTH DISEASE IN THE UNITED KINGDOM, 2001

In February 2001, the first case of foot-and-mouth disease (FMD) in 20 years was detected in a pig in a slaughterhouse in the United Kingdom. By that time the epidemic had already begun. Eleven months later, at least 6 million animals had been destroyed, including 4.9 million sheep, 700,000 cattle, 400,000 pigs, 2,000 goats, and 1,000 deer. Direct and indirect losses were calculated to exceed £12 billion (www.defra.gov.uk).[51–53]

Losses involved not only the agricultural sector of the economy but also sectors integrated with agricultural and tourism-related industries. The social costs of the outbreak and the long-term effects on farming and rural communities are incalculable.[54–58]

What was the source of the virus causing this catastrophic outbreak? And was its introduction into the United Kingdom accidental or deliberate? Today, neither question can be answered with any confidence. Why is this?

Paradoxically, more than a century ago, FMD virus (FMDV) was the first animal virus discovered and only the second viral agent described.[59] Years later it was shown to belong to the same virus family, the Picornaviruses,[60,61] as such important human pathogens as poliovirus, Hepatitis A, rhinovirus, and others. Like them, FMDV contains a positive-sense, single-stranded RNA genome of approximately 8,000 nucleotides that is directly translated into a single polyprotein upon entry into the host cell cytoplasm. The genome contains a single open reading frame (ORF) coding for a polyprotein flanked by two untranslated regions (UTRs), both predicted to display complex secondary structures.[60,61] Despite the large body of molecular knowledge concerning biochemical and genetic properties of FMDV that has accumulated over the years, fundamental questions remain unanswered about the virus and the disease.

Current estimates of FMDV genome mutation rates suggest that, on average, about one base misincorporation is likely to occur each time a single FMDV genome replicates. This should result in the introduction of every possible one-step mutation from the progenitor genotype within a single infected animal many times a day.[62] Since genomic RNA is made up of four nucleotides, a genome of 10,000 nucleotides could occupy a total of $4^{10,000}$ points in sequence space: however, only a tiny fraction of this immense number is actually realized by RNA viral genomes.[50,63] The significance of this, for forensic purposes, is that a single replicative cycle is theoretically sufficient to incorporate genetic changes in the progeny. Such a high rate of mutation makes any single-nucleotide change too common to be informative. Multiple nucleotide differences must be used for elaboration of distance matrices and phylogenetic trees.[64]

Over the last 15 years an impressive amount of sequence information from the carboxy-terminal region of the FMDV capsid protein VP1 gene has accumulated, driven in part because this region carries neutralizing antibody epitopes important for selection of suitable vaccines.[65–69] It is precisely the immunological pressure applied from the host in this area, and the resultant genetic hypervariability in the virus, that make it theoretically appropriate for analysis of phylogenetic relationships between isolates. Rates of nucleotide substitution for VP1 range from $0.5–1.5 \times 10^{-2}$ nucleotide substitutions/per site/per year (nt/s/y).[62]

Currently, approximately 2,000 partial or complete VP1 sequences of the seven FMDV serotypes are available: over 1,000 are serotype O.[70] In general, these sequences show that viral populations that are circulating and causing disease outbreaks around the world are genetically heterogeneous but related, being distributed by regions in genetically distinct virus populations known as "topotypes" that can be differentiated based on nucleotide differences of up to 15%.[65,70] These molecular methods enable scientists to unambiguously identify the strain of virus responsible for an outbreak, but do not allow precise identification of strain origin in terms of geography. In addition, the observations of unrelated and nonsystematic VP1 sequence changes provide additional confusion regarding actual phylogenetic relationships[65,67,71–74] that impairs use of VP1 phylogenetics as a forensic tool. The FMDV studies of Moya et al.[75] strongly suggest that phylogenetic reconstruction methods can infer erroneous phylogenies due to nucleotide convergences between isolates belonging to different experimental lineages that join by accident in time or space of sampling. They also point out that diverse evolutionary mechanisms acting under differing experimental dynamics generate alterations and change the frequencies of genetic variants, which can lead to the misinterpretation of the real evolutionary history.

Over the last 21 years, a strain of FMDV serotype Pan Asia O has spread from India throughout Southern Asia and the Middle East.[76] During 2000, this strain caused disease outbreaks in the Republic of Korea, Japan, Russia, Mongolia, and South Africa. In February 2001, the Pan Asia O strain spread to the U.K. Studies of sequences of the VP1-coding region from approximately 30 Pan Asia O isolates demonstrated that the U.K. virus was closely related to all of them, and nearly identical to the South African isolate O/SAR/1/2000.[77] Notably, nucleotide sequences of the VP1 coding regions of 30 Pan Asia O viruses isolated over an 11-year period differed by no more than 5%, making meaningful phylogenetic and forensic analysis extremely difficult.[78] Complete genome sequence analysis of eight Asian, African, and European isolates of Pan Asia O strains confirmed the close relationship between the South Africa and U.K. outbreaks, but failed to identify, or even imply, the mechanism of introduction or the source attribution for the latter outbreak. The results were consistent with either a common source for both the 2000 South Africa and the 2001 U.K. outbreaks or that O/SAR/1/2000 is the source of the strain that caused the U.K. outbreak.[79]

This close genetic relationship among Pan Asia O strains circulating in the world today and lack of sufficient knowledge about mutation frequency, mutation rates, mutation and recombination processes, tolerance and constraints of genes and proteins, and evolution and memory of FMD viruses in nature prevent meaningful phylogenetic and forensic analysis. Currently, results obtained lack sufficient statistical support.

This is the reason why neither of the questions posed above can be answered.

WHAT NEEDS TO BE DONE?

To advance phylogenetic hypotheses for forensic purposes, a confidence interval of 95% or greater is needed.[80] However, the knowledge to provide better statistical support, including the real-mutation range (quality of the possible substitutions) and rate of mutation of FMDV in nature, is lacking. Epidemiological, functional, and structural studies indicate that RNA viruses can tolerate restricted types and numbers of mutations at any specific time point in their evolution, suggesting the existence of critical thresholds for the expression of phenotypic traits, which may underlie *in vivo* restriction of variability rates. While high mutation and recombination rates lead to survival and adaptability of the species, variation is limited by functional and structural constraints to preserve the continuity of the “core” information.

Understanding of the principles that regulate these restrictions and the differences between *in vivo* and *in vitro* mutation tolerance is needed.[81] In other words, although evolution of RNA viruses seems to be largely unpredictable because of the nature of genetic drifts, the chance transmission bottlenecks, and the influence of host/environmental factors on the success of a determined quasispecies, there must be a finite probability that a given nucleotide and/or amino acid can be substituted and still succeed as a new master sequence of a given viral population.

To define real natural variability in FMDV isolates, databases of the globally circulating strains are needed.[109] Unfortunately, most of the molecular epidemiological studies of FMDV currently available have been limited to a single structural protein (VP1). Given the advent of high-throughput DNA sequencing, there is little reason to limit FMDV genomic analysis to a small region of VP1. In some cases, when comparing nonstructural proteins, the paucity of available sequences for comparison may explain the abundance of certain types of nucleotide and amino acid substitutions and the absence of many others, but, alternatively, it may reflect constraints on allowable variation for the viral RNA genome, probably due to the complexity of vital functions required for cell-host interactions. Clarification of the tolerance for change in different genomic regions is absolutely necessary for further progress in viral forensic analysis and interpretation.[109]

An additional problem unanswered by partial genome analysis is homologous and nonhomologous recombination and genome segment reassortment.[82–84] Recombination occurs at a high frequency in poliovirus vaccines[85,86] and appears to act as a major evolutionary force in the expansion of the human immunodeficiency viruses.[87–90] Significantly, many emerging viruses belong to virus families that actively recombine.[17,18,20] With FMDV, naturally occurring recombinant viruses have been described.[69,109] The significance of this is that classification and identification of circulating isolates based solely on VP1 may lead to highly erroneous interpretations. Using more than just VP1 sequence to characterize an FMDV isolate would address this problem.[109]

For forensic purposes, genetic profiles of viral populations should include "kits of genetic markers."[91–95] A globally representative set of informative signatures should be used. Limited FMDV sequence information from genomic regions other than VP1 has hindered the search for other genetic markers with informative value.[49] Identification of informative genetic markers will permit assessment of the probability that two or more mutations can occur within a given viral master sequence, in a similar proportion, in two distinct viral populations, providing a tool for distinguishing between two closely related individuals. The challenge now is to identify the right combination of genetic markers that cover a wide range of possible phenotypic traits. While genes with slower rates of mutation fixation may be useful for analyzing natural evolution, genes with higher rates may be useful for analysis of recent events. By careful choice of the genomic region to study, based on evolutionary divergence levels for the issue under consideration, the necessary information will be provided to find the best genetic targets for analysis of alternative hypotheses.

Analysis of complete genomic sequences of FMDV isolates will allow: (1) assessment of mutation frequency, mutation rates, mutation and recombination processes, tolerance and constraints of genes and proteins, and evolution and memory of FMD viruses in natural isolates; (2) precise identification of strains circulating around the world for attribution; (3) estimation of the probability of a given mutation occurring in nature for identification of unnaturally altered viruses and specific signatures; and (4) statistical support for hypothesis development and evolutionary model selection.

Complete genome sequence analysis of closely related FMDV isolates will better define the phylogenetic relationships between isolates, and how genetic imprints of selective struggles and bottlenecks are inherited from the parental to the next viral generation. Changes which are deleterious for the virus, sterically or physically impossible, or incompatible with the host range would not be expected to occur normally. Thus, knowledge of conserved genomic areas as well as the probability of a given mutation occurring in nature would be of great value in determining whether a virus has been unnaturally altered.

Selection of the appropriate evolutionary model is a critical premise for forensic attribution. It must be kept in mind that phylogenetic reconstructive methods involving mathematical simulations grossly simplify the relationship between organisms. The unavoidable translation of the biological properties of a virus in nature to numerical units is not an absolute and exact process. Data used in these analyses do not contain the complete historical record of the virus and are far from ideal in representing all possible evolutionary pathways of the population, which can result in erroneous phylogenies.[75] We must take into account the stochastic nature of transmission and the lack of knowledge about its consequences in viral population genetics, and the biases and gaps involuntarily inflicted during collection and processing of epidemiological data. For example, FMDV infects up to 70 species of cloven-hoofed mammals, but the extent of lesions and clinical disease are host-dependent: sheep present with much milder disease than cattle or pigs.[96] In the 2001 U.K. epidemic, sheep played a prominent role in the early stages of the outbreak, laying the groundwork for the subsequent epidemic as infection passed inadvertently from sheep to sheep. Passage of virus in sheep for an unknown number of generations has the potential to distort the shape of the evolutionary tree, possibly masking parenthood relationships with previous isolates. Thus, knowledge of genetic changes associated with replication in specific host species is needed.

Some routine procedures used for statistical evaluation in molecular epidemiology may or may not be useful for forensic purposes. For example, when the variability is too high, bootstrap support for identification of the relevant node defining the monophyletic clade becomes useless from a forensic point of view. This is because the main interest here is to find the prior probability of the two alternative hypotheses (usually prosecution and defense hypotheses). For forensic analysis, maximum likelihood testing is the method of choice for evaluating competing phylogenetic hypotheses when linked to other types of evidence (other genetic and/or epidemiological information), thus providing quantitative criteria for deciding between alternative possibilities.[80]

Statistical evaluation of alternative phylogenetic hypotheses may then be used to single out the likelihood of an isolate sharing the parental virus with other isolates as compared to the existence of a monophyletic clade that includes all the isolates related to a common source. This is a very important issue in highly polymorphic viral populations such as FMDV.

A common observation in sequence alignments is the high frequency of certain specific nucleotide and amino acid substitutions.[27,75,97–100] For example, during the course of viral evolution, variable positions in capsid proteins of FMDV are alternatively occupied by a small subset of all possible amino acids, and a true accumulation of amino acid substitutions is not observed.[101] Alignment of the VP1 nucleotide sequences of recent Pan Asia O isolates showed that the most closely related virus to the U.K. 2001 strain was the South Africa 2000 strain, which differs in just one nucleotide in VP1, and the Japan 2000 isolate, which differs by an additional three nucleotides. It is impossible to define a chronological link between these isolates other than that of the monophyletic inclusion of all isolates with less than 5% differences between each other as originating from a common ancestor.[79] Molecular epidemiology in a forensic context should give priority to individual evaluation of pertinence to a specified group or individual. Identification of less-variable genomic areas with high informative content would be of great interest for application of current phylogenetic procedures.

Finally, new algorithms are needed for forensic investigations to examine the cladistic structure of trees as a function of the distance of each case from the first infection source, and to evaluate the precision and robustness of the historical reconstruction.[52,89,102] An estimate of the history of transmission events in a particular epidemic can be extracted from reconstructed “epidemic trees” in which the case-reproduction ratio for the spread of the infectious disease has statistical reliability.[103]

In summary, development of viral forensics as a scientific discipline will need to combine expertise in conventional forensics with knowledge of viral genomics, phylogenetics, and informatics. Acceptability of current molecular phylogenetic analysis is now hampered by a lack of statistical confidence in the conclusions.[104–107] Methods that allow a statistical evaluation of the likelihood of a virus belonging to a specific lineage must be developed.[108] This will require genomic databases of currently circulating viral strains and identification of the most informative viral signatures.

REFERENCES

1. Larkin M. Microbial forensics aims to link pathogen, crime, and perpetrator. *Lancet Infect Dis* 2003;3:180.

2. Popovic T, Glass M. Laboratory aspects of bioterrorism-related anthrax—From identification to molecular subtyping to microbial forensics. *Croat Med J* 2003;44:336–341.
3. Meng Z, Zaykin DV, Xu CF, Wagner M, Ehm MG. Selection of genetic markers for association analyses, using linkage disequilibrium and haplotypes. *Am J Hum Genet* 2003;73:115–130.
4. Latorra D, Campbell K, Wolter A, Hurley JM. Enhanced allele-specific PCR discrimination in SNP genotyping using 3' locked nucleic acid (LNA) primers. *Hum Mutat* 2003;1:79–85.
5. Kwok PY, Chen X. Detection of single nucleotide polymorphisms. *Curr Issues Mol Biol* 2003;5:43–60.
6. Kaderali L, Deshpande A, Nolan JP, White PS. Primer-design for multiplexed genotyping. *Nucleic Acids Res* 2003;31:1796–1802.
7. Hauser J, Leszczynska A, Samochowiec J, Czerski PM, Ostapowicz A, Chlopocka M, Horodnicki J, Rybakowski JK. Association analysis of the insertion/deletion polymorphism in serotonin transporter gene in patients with affective disorder. *Eur Psychiatry* 2003;18:129–132.
8. Perez-Lezaun A, Calafell F, Seielstad M, Mateu E, Comas D, Bosch E, Bertranpetit J. Population genetics of Y-chromosome short tandem repeats in humans. *J Mol Evol* 1997;45:265–270.
9. Hatzaki A, Loukopoulos D, Spiliopoulou H, Koutselinis A. A study of five variable number tandem repeat (VNTR) loci in the Greek population. *Mol Cell Probes* 1995;9:129–133.
10. Jin L, Chakraborty R. Population structure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics. *Heredity* 1995;74:274–285.
11. Dallapiccola B, Novelli G, Spinella A. PCR DNA typing for forensics. *Nature* 1991;354:179.
12. Wen SY, Wang H, Sun OJ, Wang SQ. Rapid detection of the known SNPs of CYP2C9 using oligonucleotide microarray. *World J Gastroenterol* 2003;9:1342–1346.
13. Ahmed N, Alam M, Abdul Majeed A, Asad Rahman S, Cataldi A, Cousins D, Hasnain SE. Genome sequence based, comparative analysis of the fluorescent amplified fragment length polymorphisms (FAFLP) of tubercle bacilli from seals provides molecular evidence for a new species within the Mycobacterium tuberculosis complex. *Infect Genet Evol* 2003;2:193–199.
14. Ye J, Parra EJ, Sosnoski DM, Hiester K, Underhill PA, Shriver MD. Melting curve SNP (McSNP) genotyping: A useful approach for diallelic genotyping in forensic science. *J Forensic Sci* 2002;47:593–600.
15. Broude NE, Driscoll K, Cantor CR. High-level multiplex DNA amplification. *Antisense Nucleic Acid Drug Dev* 2001;11:327–332.
16. Murphy FA. Virus Taxonomy. *In* Fields BN, Knipe DM, Howley PM, Chanock RM, Melnick JL (eds). *Fields Virology.* Philadelphia, PA: Lippincott-Raven, 1996;15–27.
17. Morse SS. The viruses of the future? Emerging viruses and evolution. *In* Morse SS (ed). *The Evolutionary Biology of Viruses.* New York, NY: Raven Press, 1994;25–335.
18. Lederberg J, Shope RE, Oaks SC, eds. *Emerging Diseases: Microbial Threats to Health in the United States.* Washington, DC: Nat Acad Press, 1992.
19. Lopez N, Padula P, Rossi C, Lazaro ME, Franze-Fernandez MT. Genetic identification of a new Hantavirus causing severe pulmonary syndrome in Argentina. *Virology* 1996;220:223–226.
20. Murphy FA, Nathanson N. The emergence of new virus diseases—An overview. *Semin Virol* 1994;5:87–102.
21. Nichol ST, Spiropoulou CF, Morzunov S, Rollin PE, Ksiazek TG, et al. Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. *Science* 1993;262:914–917.
22. Melson H. USA: Mysterious respiratory illness. *Lancet* 1993;341:1526.

23. Anderson JF, Andreadis TG, Vossbrinck CR, Tirrell S, Wakem EM, French RA, Garmendia AE, Van Kruiningen HJ. Isolation of West Nile virus from mosquitoes, crows, and a Cooper's hawk in Connecticut. *Science* 1999;286:2331–2333.
24. Update: Outbreak of severe acute respiratory syndrome—Worldwide 2003. *Morb Mortal Wkly Rep* 2003;28:241–248.
25. Friedberg EC, Walker GC, Siede W. *DNA repair and mutagenesis*. Washington, DC: ASM Press, 1995.
26. Steinhauer D, Domingo E, Holland JJ. Lack of evidence of proofreading mechanisms associated with an RNA virus Polymerase. *Gene* 1992;122:281–288.
27. Modrich P, Lahue R. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Ann Rev Biochem* 1996;65:101–133.
28. Domingo E, Holland JJ, Ahlquist P. *RNA Genetics*, Vols. 2, 3. Boca Raton, FL: CRC Press, 1988.
29. Domingo E, Martinez-Salas E, Sobrino F, de la Torre JC, Portela A, et al. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: Biological relevance—A review. *Gene* 1985;40:1–8.
30. Domingo E, Sabo DL, Taniguchi T, Weissmann C. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 1978;13:735–744.
31. Holland JJ, Spindler K, Horodyski F, Grabau E, Nichol S. Rapid evolution of RNA genomes. *Science* 1982;215:1577–1585.
32. Domingo E, Holland JJ. Mutation rates and rapid evolution of RNA viruses. *In* Morse SS (ed). *Evolutionary Biology of Viruses*. New York, NY: Raven Press, 1994;161–184.
33. Drake JW. Rates of spontaneous mutations among RNA viruses. *Proc Nat Acad Sci* USA 1993; 90:4171–4175.
34. Drake JW, Holland JJ. Mutation rates among RNA viruses. *Proc Nat Acad Sci* USA 1999; 96:13910–13913.
35. Temin HM. The high rate of retrovirus variation results in rapid evolution. *In* Morse SS (ed). *Emerging Viruses*. Oxford: Oxford Univ Press, 1993;219–225.
36. Wain-Hobson S. Is antigenic variation of HIV important for AIDS and what might be expected in the future? *In* Morse SS (ed). *Evolutionary Biology of Viruses*. New York, NY: Raven Press, 1994;185–209.
37. Wain-Hobson S. 1996. Running the gamut of retroviral variation. *Trends Microbiol* 1996;4:135–141.
38. Holland JJ, de la Torre JC, Steinhauer D. RNA virus populations as quasispecies. *Curr Top Microbiol Immunol* 1992;176:1–20.
39. Domingo E, Webster RG, Holland JJ. *Origin and Evolution of Viruses*. San Diego, CA: Academic Press, 1999.
40. Eigen M. *Steps Towards Life*. Oxford: Oxford University Press, 1992.
41. Eigen M, Biebricher CK. Sequence space and quasispecies distribution. *In* Domingo E, Holland JJ, Alquist P (eds). *RNA Genetics*, vol. 3. Boca Raton, FL: CRC Press, 1998; 211–245.
42. Gell-Mann M. Complex adaptive systems. *In* Cowan GA, Pines D, Meltzer D (eds). *Complexity, Metamorphosis, Models and Reality*. Reading, MA: Addison Wesley Publishing, 1994; 17–45.
43. Eigen M. Self organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 1971;58:465–523.
44. Eigen M, Schuster P. *The Hypercycle. A Principle of Natural Self-organization*. Berlin: Springer Verlag, 1979.
45. Batschelet E, Domingo E, Weissmann C. The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. *Gene* 1976;1:27–32.

46. Domingo E, Holland JJ. Mutation rates and rapid evolution of RNA viruses. *In* Morse SS (ed). *Evolutionary Biology of Viruses*. New York, NY: Raven Press, 1994;161–184.
47. Domingo E, Sabo D, Taniguchi T, Weissmann C. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 1978;13:735–744.
48. Domingo E, Webster RG, Holland JJ. *Origin and Evolution of Viruses*. San Diego, CA: Academic Press, 1999.
49. Mohapatra JK, Sanyal A, Hemadri D, Tosh C, Sabarinath GP, Venkataramanan R. Sequence and phylogenetic analysis of the L and VP1 genes of foot-and-mouth disease virus serotype Asia 1. *Virus Res* 2002;87:107–118.
50. Eigen M, Biebricher CK. Sequence space and quasispecies distribution. *In* Domingo E, Holland JJ, Alquist P (eds). *RNA Genetics*, vol. 3. Boca Raton, FL: CRC Press, 1988;211–245.
51. Editorial. It has been a difficult year for the British Ministry of Agriculture. *Nature Immunol* 2001;2:565.
52. Martin MJ, Gonzalez-Candelas F, Sobrino F, Dopazo J. A method for determining the position and size of optimal sequence regions for phylogenetic analysis. *J Mol Evol* 1995;41:1128–1138.
53. Aggarwal N, Zhang Z, Cox S, Statham R, Alexandersen S, Kitching RP, Barnett PV. Experimental studies with foot-and-mouth disease virus, strain O, responsible for the 2001 epidemic in the United Kingdom. *Vaccine* 2002;20:2508–2515.
54. Thompson D, Muriel P, Russell D, Osborne P, Bromley A, Rowland M, Creigh-Tyte S, Brown C. Economic costs of the foot and mouth disease outbreak in the United Kingdom in 2001. *Rev Sci Tech* 2002;21:675–687.
55. The Royal Society of Edinburgh. Inquiry into foot and mouth disease in Scotland. 2002. (www.ma.hw.ac.uc)
56. The Royal Society. Infectious diseases of livestock. Summary and recommendations. 2002. (www.royalsoc.ac.uk)
57. Foot and mouth disease 2001: Lessons to be learned inquiry. House of Commons 2002. (archive.cabinet-office.gov)
58. Origin of the Foot and Mouth Disease epidemic in 2001, Department of Environment, Food and Rural Affairs, 2002 and 2003. (www.defra.gov.uk)
59. Loeffler F, Frosh P. Berichte der Kommission zur Erforschung der Maul-und Klauenseuche bei dem Institut fuer Infektioneskrankheiten in Berlin. Zbl Bakter 1898;Abt 1;23, 371–391.
60. Rueckert RR, Melnick JL, Couch RB, Blaine-Hollinger F, Ticehurst JR. Picornaviridae. *In* Fields BN, Knipe DM, Howley PM (eds). *Fields Virology*, vol 1. Philadelphia, PA: Lippincott-Raven Publishers, 1995;609–810.
61. Gromeier M, Wimmer E, Gorbalenya AE. Genetics, pathogenesis and evolution of Picornaviruses. *In Origin and Evolution of Picornaviruses*. Academic Press, 1999;287–343.
62. Haydon DT, Samuel AR, Knowles NJ. The generation and persistence of genetic variation in foot-and-mouth disease virus. *Prev Vet Med* 2001;51:111–124.
63. Eigen M. The origin of genetic information: viruses as models. *Gene* 1993;135:37–47.
64. Hall BG. *Phylogenetic Trees Made Easy: A How-to Manual for Molecular Biologists*. Boston, MA: Sinauer Associates, 2001;197.
65. Knowles NJ, Samuel AR. Molecular epidemiology of foot-and-mouth disease virus. *Virus Res* 2003;91:65–80.
66. Feng Q, Chen X, Ma O, Liu Y, Ding M, Collins RA, Ko LS, Xing J, Lau LT, Yu AC, Chen J. Serotype and VP1 gene sequence of a foot-and-mouth disease virus from Hong Kong (2002). *Biochem Biophys Res Comm* 2003;302:715–721.
67. Tosh C, Hemadri D, Sanyal A, Bandyopadhyay SK. Genetic and antigenic analysis of two recently circulating genotypes of type A foot-and-mouth disease virus in India: evidence for positive selection in the capsid-coding genes. *Archs Virol* 2003;148:853–869.

68. Singh M, Mohan BM, Suryanarayana VV. Serological and molecular analysis of serotype O foot-and-mouth disease virus isolated from disease outbreaks in India during 1987–91. *Virus Res* 1996;43:45–55.
69. Nunez JI, Martin MJ, Piccone ME, Carrillo E, Palma EL, Dopazo J, Sobrino F. Identification of optimal regions for phylogenetic studies on VP1 gene of foot-and-mouth disease virus: Analysis of types A and O Argentinian viruses. *Vet Res* 2001;32:31–45.
70. Samuel AR, Knowles NJ. Foot and mouth disease type O viruses exhibit genetically and geographically distinct evolutionary lineages (topotypes). *J Gen Virol* 2001;82:609–621.
71. Tosh C, Hemadri D, Sanyal A. Evidence of recombination in the capsid-coding region of type A foot-and-mouth disease virus. *J Gen Virol* 2002;83:2455–2460.
72. Nayak B, Pattnaik B, Tosh C, Sanyal A, Hemadri D, Patil SS, Venkataramanan R. Genetic and antigenic analysis of type A foot-and-mouth disease viruses isolated in India during 1987–1996. *Acta Virol* 2001;45:13–21.
73. Beck E, Strohmaier K. Subtyping of European foot-and-mouth disease virus strains by nucleotide sequence determination. *J Virol* 1987;61:1621–1629.
74. Kweon CH, Ko YJ, Kim WI, Kwon BJ, Hyun BH, Sohn HJ, Choi KS, Shin JH. Molecular characterization of foot-and-mouth disease virus O/SKR/2000. *Virus Res* 2002;90:15–22.
75. Fares MA, Barrio E, Becerra N, Escarmis C, Domingo E, Moya A. The foot-and-mouth disease RNA virus as a model in experimental phylogenetics. *Int Microbiol* 1998;1:311–318.
76. Hemadri D, Tosh C, Sanyal A, Venkataramanan R. Emergence of a new strain of type O foot-and-mouth disease virus: Its phylogenetic and evolutionary relationship with the PanAsia pandemic strain. *Virus Genes* 2002;25:23–34.
77. Knowles NJ, Samuel AR, Davies PR, Kitching RP, Donaldson AI. Outbreak of foot-and-mouth disease virus serotype O in the UK caused by a pandemic strain. *Vet Rec* 2001;48:258–259.
78. Samuel AR, Knowles NJ, Mackay DK. Genetic analysis of type O viruses responsible for epidemics of foot-and-mouth disease in North Africa. *Epidemiol Infect* 1999;122:529–538.
79. Mason PW, Pacheco JM, Zhao QZ, Knowles NJ. Comparisons of the complete genomes of Asian, African and European isolates of a recent foot-and-mouth disease virus type O pandemic strain (PanAsia). *J Gen Virol* 2003;84:1583–1593.
80. Gonzalez-Candelas F, Bracho MA, Moya A. Molecular epidemiology and forensic genetics: application to a hepatitis C virus transmission event at a hemodialysis unit. *J Infect Dis* 2003;187:352–358.
81. Carrillo C, Plana J, Mascarella R, Bergada J, Sobrino F. Genetic and phenotypic variability during replication of foot-and-mouth disease virus in swine. *Virology* 1990;179:890–892.
82. Lai MMC. Genetic recombination in RNA viruses. *Curr Top Microbiol Immunol* 1992; 176:21–32.
83. Simmon AE, Bujarski JJ. RNA-RNA recombination and evolution in virus-infected plants. *Ann Rev Phytopathol* 1994;32:337–362.
84. Strauss JH. Recombination in the evolution of RNA viruses. *In* Morse SS (ed). *Emerging Viruses*. Oxford: Oxford University Press, 1993;241–251.
85. Minor PD. The molecular biology of poliovaccines. *J Gen Virol* 1992;73:3065–3077.
86. Wimmer E, Hellen CUT, Cao X. Genetics of poliovirus. *Ann Rev Genet* 1993;27:353–436.
87. Robertson DK, Sharp PM, McCutchan FE, Hahn BH. Recombination in HIV-1. *Nature* 1995; 374:124–126.
88. Sharp PM, Robertson DL, Gao F, Hahn BH. Origin and diversity of human immunodeficiency viruses. *AIDS* 1994;8:S27–S42.
89. Martin NJ, Nunez JI, Sobrino F, Dopazo J. A procedure for detecting selection in highly variable viral genomes: Evidence of positive selection in antigenic regions of capsid protein VP1 of foot-and-mouth disease virus. *J Virol Methods* 1998;74:215–221.
90. Bailes E, Gao F, Bibollet-Ruche F, Courgnaud V, Peeters M, Marx PA, Hahn BH, Sharp P. Hybrid origin of SIV in chimpanzees. *Science* 2003;300:1713.

91. Krane DE, Allen RW, Sawyer SA, Petrov DA, Hartl DL. Genetic differences at four DNA typing loci in Finnish, Italian, and mixed Caucasian populations. *Proc Nat Acad Sci* USA 1992; 89:10583–10587.
92. Devlin B, Risch N. Ethnic differentiation at VNTR loci, with special reference to forensic applications. *Am J Hum Genet* 1992;51:534–548.
93. Goncalves R, Jesus J, Fernandes AT, Brehm A. Genetic profile of a multi-ethnic population from Guinea-Bissau (West African coast) using the new PowerPlex 16 system kit. *Forensic Sci Int* 2002;129:78–80.
94. Agrawal S, Muller B, Bharadwaj U, Bhatnagar S, Khan F. Evaluation of six short tandem repeat loci in forensics: North Indian populations. *J Forensic Sci* 2002;47:686–689.
95. Pattnaik B, Venkataramanan R, Tosh C, Sanyal A, Hemadri D, Samuel AR, Knowles NJ, Kitching RP. Genetic heterogeneity of Indian field isolates of foot-and-mouth disease virus serotype O as revealed by partial sequencing of 1D gene. *Virus Res* 1998;55:115–127.
96. Pereira HG. Foot and mouth disease. *In* Gibbs EPJ (ed). *Virus Diseases of Food Animals.* London: Academic Press, 1981;333–363.
97. Joyce CM, Steitz TA. Function and structure relationships in DNA polymerases. *Ann Rev Biochem* 1994;63:777–822.
98. Couderc T, Guedo N, Calvez V, Pelletiner I, Hogle J, et al. Substitutions in the capsids of poliovirus mutants selected in human neuroblastoma cells confer on the Mahoney type 1 strain a phenotype neurovirulent in mice. *J Virol* 1994;68:8386–8391.
99. De la Torre JC, Giachetti C, Semler BL, Holland JJ. High frequency of single-base transitions and extreme frequency of precise multiple-base reversion mutations in poliovirus. *Proc Nat Acad Sci* USA 1992;89:2531–2535.
100. Lu Z, Rezapkin GV, Douthitt MP, Ran Y, Asher DM, et al. Limited genetic changes in the Sabin I strain poliovirus occurring in the central nervous system of monkeys. *J Gen Virol* 1996;77:273–280.
101. Martinez MA, Dopazo J, Hernandez J, Mateu MG, Sobrino F, et al. Evolution of the capsid protein genes of foot-and mouth disease virus: Antigenic variation without accumulation of amino acid substitutions over six decades. *J Virol* 1992;66:3557–3565.
102. Qinones-Mateu ME, Holaguin A, Dopazo J, Najera I, Domingo E. Point mutant frequencies in the pol gene of human immunodeficiency virus type 1 are two- to threefold lower than those of env. *AIDS Res Hum Retroviruses* 1996;12:1117–1128.
103. Haydon DT, Chase-Topping M, Shaw DJ, Matthews L, Friar JK, Wilesmith J, Woolhouse ME. The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proc R Soc Lond B Biol Sci* 2003;270:121–127.
104. Ou CY, Ciesielki CA, Myers G, et al. Molecular epidemiology of HIV transmission in a dental practice. *Science* 1992;256:1165–1171.
105. Power JP, Lawlor E, Davidson F, Holmes EC, Yap PL, Simmonds P. Molecular epidemiology of an outbreak of infection with hepatitis C virus in recipients of anti-D immunoglobulin. *Lancet* 1995;345:1211–1213.
106. Esteban JI, Martell M, Carman W, Gomez J. The impact of rapid evolution of the hepatitis viruses. *In* Domingo E, Holland JJ, Webster JR (eds). *Origin and Evolution of the Viruses.* San Diego, CA: Academic Press, 1999;345–376.
107. Izopet J, Pasquier C, Sanders K, Puel J, Rostaing L. Molecular evidence for nosocomial transmission of hepatitis C virus in a French hemodialysis unit. *J Med Virol* 1999; 58(2):139–144.
108. Hillis DM, Huelsenbeck JP, Cunningham CW. Application and accuracy of molecular phylogenies. *Science* 1994;264:671.
109. Carrillo C, Tulman ER, Delhon G, Lu Z, Carreno A, Vagnozzi A, Kutish GF, Rock DL. Comparative genomics of Foot-and-Mouth disease virus. *J Virol* 2005, *in press*.

Investigation of Suspicious Disease Outbreaks

LYNDA COLLINS KELLEY
U.S. Department of Agriculture, Food Safety and Inspection Service, Office of Public Health Science, Athens, Georgia

ROGER G. BREEZE
Centaur Science Group, Washington, DC

Most of the terrorism experienced globally has involved highly localized attacks with explosives or toxic chemicals that produce mass casualties at the time of the event or soon after. Such incidents are obvious, and the extent of the damage can be readily and immediately perceived. This chapter is dedicated to investigation of covert attacks with biological weapons, in which signs and symptoms of clinical illness are usually delayed so that the scope of the problem grows over hours and days. Without physical evidence of an attack or a credible announcement, the deliberate nature of the incident may not be suspected or discovered for some time, if ever. Given the heightened national state of alert currently, almost all outbreaks of disease, regardless of size, should be evaluated considering the possibility of bioterrorism or agroterrorism. In many cases, a rapid, well conducted etiologic and epidemiologic investigation paired with an analysis of the distribution and number of reported cases will provide important clues regarding the source of infection. Given the threat of biological weapons use, we must develop criteria and data to differentiate natural and deliberate disease outbreaks.

NATURAL AND DELIBERATE DISEASE

Differentiating natural and deliberate outbreaks of disease is important not only for the determination of criminal events within countries, but is also central to building international confidence that unusual outbreaks are not the result of prohibited weapons-related activities. The Biological and Toxin Weapons Convention of 1972[1] prohibits development, production, stockpil-

Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

ing, or acquisition of biological or toxin weapons: the intent to prohibit use of biological weapons was reinforced at the Fourth Review Conference.[2] The Biological Weapons Statute, Chapter 10 of Title 18, United States Code, was amended by the U.S.A. Patriot Act of 2001[3] to criminalize not only the use of a biological weapon but also possession of a biological agent, toxin, or delivery system . . . that is not reasonably justified by a prophylactic, protective, . . . research or other peaceful purpose. Unlike nuclear and most chemical weapons, the dual-use nature of the pathogens that might be employed as biological weapons (BWs) makes definitive inspections of facilities for pathogens and prohibited weapons-related activities difficult and at times nearly impossible. There is international agreement that investigation of disease outbreaks will most likely prove to be much more valuable than facility inspections in determining that a biological weapon has been used or accidentally released. At the Second Review Conference in 1986,[4] it was agreed that "States Parties are to implement, on the basis of mutual cooperation, the following measures, in order to prevent or reduce the occurrence of ambiguities, doubts and suspicions, and in order to improve international cooperation in the field of peaceful bacteriological (biological) activities: Exchange of information on all outbreaks of infectious diseases and similar occurrences caused by toxins that seem to deviate from the normal pattern as regards type, development, place, or time of occurrence."

At an Ad Hoc meeting of Experts on the Biological Weapons Convention Treaty in 1987,[5] States Parties were recommended to take guidance in determining what constitutes an outbreak from the World Health Organization (WHO) and from the Dictionary of Epidemiology. The term "epidemic" has evolved to apply to any disease or event that occurs in clear excess of normal expectancy.[6] It was also agreed that since no universal standards exist for what might constitute a deviation from the normal pattern, States Parties were encouraged to utilize reporting systems within WHO. States Parties were also to provide background information (reporting) on diseases caused by organisms which met the criteria for risk groups III and IV [according to the classification in the WHO Laboratory BioSafety Manual],[7] even though the occurrence of these diseases does not necessarily constitute a deviation from normal patterns. Unfortunately, annual declarations made by States Parties under this confidence-building measure have been incomplete and are therefore not as useful for establishing a baseline of expected normal occurrence of disease as would be desired.[8]

Potential criteria for considering an outbreak unusual were developed by the Ad Hoc Group in 1998:[9]

a) The disease is being reported for the first time in the region and was never endemic;

b) The epidemic occurs outside its normal anticipated season;

c) The reservoir host and/or insect vector of the disease do not occur in or were previously eradicated from the affected region;

d) The disease appears to be transmitted by an uncommon or unusual route;

e) The epidemiological features of the disease suggest increased virulence of the organism manifested in the form of increased case fatality rate;

f) The causative agent has a higher survival time even in adverse environmental conditions and shows unusual resistance;

g) The causative agent is capable of establishing new natural reservoirs to facilitate continuous transmission;

h) Outbreaks of the disease in a population with a high level of immunity due to vaccination suggest that the causative agent is modified;

i) The disease is caused by an agent with an unusual population subset or in an unexpected age group;

j) The epidemiology of the disease suggests an abnormal reduction in the incubation period of the disease;

k) The epidemiology of the outbreak strongly points to an etiology, but isolation and identification of the suspected agent is not possible by established means;

l) The characteristics of the causative agent differ from the known characteristics of that agent prevalent in the territory of the State Party. (Characteristics could include strain type, sequence, antibiotic resistance pattern, etc.)

Another factor that might indicate that an outbreak is outside the normal realm of disease would be occurrence of an exceptionally large epidemic, especially if patients presented with fever and severe respiratory or gastrointestinal complaints (suggesting biological agents), or respiratory, ocular, cutaneous, or neurologic symptoms (suggesting toxin and/or chemical exposure).[10] However, it should be noted that there have been extremely large natural outbreaks of disease caused by *Salmonella* in ice cream and *Cryptosporidia* in drinking water, among others.

Patients who have a relatively uncommon disease caused by a pathogen with BW potential merit particular attention. Pathogens and toxins of concern are identified in the United Nations Special Commission (UNSCOM) Biological Agents List, which is based on the WHO manual,[7] the Australia Group Lists,[11–13] and the U.S. Select Agent List.[14,15] Disease outbreaks of unknown cause should be treated in the same way as those related to listed pathogens of concern. But again, it should be borne in mind that several documented biocrimes have employed common foodborne pathogens such as Salmonella and Shigella.[6–18]

Other situations that should arouse suspicion would be: epidemics impacting a discrete population, many cases of unexplained disease or death, more severe disease than is usually expected for a specific pathogen, failure to respond to standard therapy, unusual routes of exposure, disease normally transmitted by a vector that is not present in the local area, multiple simultaneous or serial epidemics of different diseases in the same population, or a single infection with an uncommon agent (smallpox, some viral hemorrhagic fevers).[19]

Other obvious indications might include: intelligence information that an adversary has access to a biological agent; claimed responsibility for an event; discovery of associated production equipment, dispersal devices or munitions; and evidence of product tampering. Suspicion might also arise from the presence of a similar genetic type among agents isolated from distinct sources at different times or locations, higher attack rates in those exposed in certain areas (such as inside a building if released indoors), and outbreaks of the same illness simultaneously in noncontiguous areas.[9,19] Presence of multiple strains of a single pathogen within one clinical sample has also been used to document an unnatural outbreak.[20] Occurrence of pathogens in an uncommon food vehicle for that pathogen[17] and contamination of multiple food items with the same pathogen[16] also indicate potential nefarious activity.

IMPORTANCE OF DISEASE SURVEILLANCE

The best tool to help differentiate an intentional attack from a natural epidemic is a robust disease surveillance and reporting system. Recognition and understanding of a deviant event are greatly compromised until the patterns of natural disease outbreaks are known. Developed countries generally have a historical record of natural disease outbreaks sufficient to form a distinct background against which potential covert acts can be judged. This is not true for countries that lack adequate disease surveillance systems because the normal baseline of endemic diseases is not known. The better the public health and agricultural health systems in place around the world, and the better and more complete the reporting of natural disease outbreaks, the less likely it is that an attack with biological weapons could be carried out covertly.[21]

In the International Animal Health Code of the International Office of Epizootics[22] (OIE, now changing its name to the World Animal Health Organisation), disease surveillance is defined as the continuous investigation of a given population to detect the occurrence of disease for control purposes, which may involve testing of a part of the population. Surveillance involves collection of important and essential information on all significant factors of disease outbreaks that allow us to learn how and why the disease has occurred, and whether the disease is changing, spreading into new areas, or infecting new species. The art of disease surveillance was well described by Dorothy Preslar:[23]

"When some people talk about national disease surveillance they often appear to be talking only about those activities distinguished above as monitoring. By whatever name it is known, it is an important function. But it is like gauging the size of an iceberg by looking at its tip. Surveillance, on the other hand, measures the total mass of the iceberg, plus (to stretch the analogy a bit) it explores the water around it, the air above it, the ice mass it broke away from, the rate of its melt, the direction it moves and at what speed, its relationship to other icebergs in the same area. As applied to outbreaks of animal disease, it is an investigation that measures the possibilities existing in individual herds and flocks, in breeding operations, in feed lots, in imports, in vaccination programs, in feed stuffs, in climate anomalies, and even in the disposal of aborted fetuses and manure. It takes time, it takes talent, it takes money and it takes enormous prescience to know where to deploy those resources to the best advantage."

Good surveillance systems routinely monitor changes in disease patterns as well as document and understand factors that have initiated the change. Effective surveillance systems have timely dissemination of that information to decision makers. Thorough understanding of the normal occurrence of diseases is imperative for informed investigation of suspicious outbreaks.

GLOBAL DISEASE REPORTING SYSTEMS

There are several excellent web-based programs for international disease surveillance and for reporting of human and animal diseases. Occurrence of specific human diseases can be searched by country through the WHO at http://www.who.int/en/. The WHO Global Outbreak Alert and Response Network provides rapid information on identification, confirmation, and investigation of human diseases of interest throughout the world at http://www.who.int/csr/outbreaknetwork/en/. The OIE maintains emergency alerts, weekly, monthly, and yearly reports, and disease archives for all participating countries at http://www.oie.int/eng/info/hebdo/A_INFO.HTM. Global animal health disease reporting and surveillance data from the Federation of American Scientists can be found at http://www.fas.org/ahead/outbreak.htm. The United Nations Food and Agriculture Organization Emergency Prevention System (EMPRES) for Transboundary Animal and Plant Pests and Diseases provides emergency alerts for high-impact animal diseases at http://www.fao.org/ag/search/agfind.asp: there are plans to add disease mapping. A unique resource for near real-time reporting of disease outbreaks or suspected incidents is maintained by ProMed Mail (http://www.fas.org/promed/), which is also supported by the Federation of American Scientists. ProMed allows e-mail posting of disease occurrence (after in-house expert review) and subsequent comment by disease specialists around the world, as well as frequent updates as additional information arises. ProMed provides early notification in that reports are often submitted without waiting for confirmation by authorized national agencies, which can be advantageous. However, information obtained through ProMed should be matched with subsequent laboratory results to validate the accuracy of the reporting.

There are also reporting systems for specific diseases, such as West Nile virus maps and statistics at http://npic.orst.edu/wnv/mapstatistics.htm; anthrax data at http://www.vetmed.lsu.edu/whocc/mp_world.htm, and the WHO world influenza monthly map at http://rhone.b3e.jussieu.fr/flunet/www/f_recent_activity.html.

Finally, the Global Emerging Infections Sentinel Network (GeoSentinel) consists of travel and tropical medicine clinics around the world that monitor geographic and temporal trends in morbidity among travelers and other globally mobile populations.[24,25] At http://www.istm.org/geosentinel/geosentinel_main.html, a rapid worldwide query and response function electronically links 1,500 International Society of Travel Medicine providers around the world.

In addition to formal surveillance systems, outbreaks are detected by hospital data (large numbers of cases being admitted), individual laboratory data, and reports in the media. In countries where no disease surveillance systems exist, the media are an important source of information. Although most countries report animal and plant diseases of international economic significance, some are still reluctant to submit information because of the negative impact that a positive report will have on trade: the same can be said of human diseases that impact tourism.

U.S. DISEASE SURVEILLANCE AND REPORTING SYSTEMS

The Centers for Disease Control and Prevention (CDC) receives weekly mortality reports from 122 U.S. cities and metropolitan areas: these summarize the total number of deaths occurring each week, as well as the number due to pneumonia and influenza. Reports are published in the Morbidity and Mortality Weekly Report (MMWR) at http://www.cdc.gov/epo/dphsi/phs.htm#121. The National Notifiable Diseases Surveillance System (NNDSS), maintained by the CDC Epidemiology Program Office, is a mechanism for regular collection, compilation, and publication of reports of diseases that are notifiable at the national level.[26,27] Data on selected notifiable infectious diseases are published weekly in the MMWR and at year-end in the *Summary of Notifiable Diseases, United States* (at http://www.cdc.gov/epo/dphsi/phs.htm). The Public Health

Laboratory Information System (PHLIS) collects data on cases and/or isolates of specific notifiable infections from every state: reports can be found at http://www.cdc.gov/ncidod/dbmd/phlisdata/default.htm.[28]

The CDC also supports an Unexplained Deaths and Critical Illnesses Surveillance System to improve understanding of the causes of specific infectious disease syndromes for which an etiologic agent is frequently not identified.[29] National and international surveillance data for clusters of unexplained deaths and illnesses can be found at: http://www.cdc.gov/ncidod/dbmd/diseaseinfo/unexplaineddeaths_t.htm.

Animal disease surveillance in the U.S. is primarily state-based. Surveillance of endemic diseases among farm livestock is facilitated by one or more of the following: the state Animal Disease Diagnostic Laboratory (often a part of the state's Land Grant university), the state Cooperative Extension Service, industry laboratories, and private veterinarians who report to the state veterinary medical officer and state public health veterinarian. Suspicion or preliminary diagnosis of an OIE List A disease mandates immediate reporting to the U.S. Department of Agriculture's Animal and Plant Health Inspection Service (USDA APHIS).

APHIS's Veterinary Services (http://www.aphis.usda.gov/vs/index.html) and Centers of Epidemiology (http://www.aphis.usda.gov/vs/ceah/cnahs/nsu/EmergingSurv.htm) monitor certain diseases of domestic livestock, but do not survey pets, wildlife, or exotic animals. APHIS has an emerging animal diseases syndromic surveillance program, EVE (Emerging Veterinary Event) that reports unusual animal health incidents. Syndromic surveillance is the monitoring of the frequency of illnesses with a specified set of clinical features in a given population without regard to a specific diagnosis. Syndromic surveillance can give an early indication that an unusual event is occurring before clinical or laboratory confirmations are conducted. Additional livestock disease surveillance information is obtained through USDA's Food Safety and Inspection Services' slaughter inspection of cattle, swine, sheep, goats, and poultry intended for human consumption.

The U.S. Geological Survey's National Wildlife Health Center maintains wildlife disease information at http://wildlifedisease.nbii.gov.

Unfortunately, no single federal agency is responsible for reporting of all animal diseases. The need for a centralized reporting mechanism for animal diseases was well demonstrated in the West Nile outbreak of 1999[30,31] and the 2003 monkeypox introduction in the Midwest.[32]

Surveillance for plant diseases, pests, and invasive species is a cooperative state and federal effort. The Federal organization is APHIS's Plant Protection and Quarantine (PPQ) section (http://www.aphis.usda.gov/ppq/). In some states the cooperating partner is the state agriculture department, in others, the Land Grant University. The Cooperative Agricultural Pest Survey is a combined federal and state effort to conduct surveillance, detection, and monitoring of agricultural crop pests and biological control agents, including weeds, plant pathogens, insects, nematodes, and other invertebrate organisms. The survey maintains a National Agricultural Pest Information System database (NAPIS) at http://ceris.purdue.edu/napis: this receives information from a federally sponsored initiative, the National Plant Diagnostic Network, presently composed of centers in California, Florida, Kansas, Michigan, and New York.

Several surveillance systems report foodborne outbreaks. The Foodborne Diseases Active Surveillance Network (FoodNet) is a collaborative project between CDC, USDA, and the U.S. Food and Drug Administration (FDA).[33,34] FoodNet has active surveillance for foodborne diseases and related epidemiologic studies designed to help public health officials better understand the epidemiology of disease (at http://www.cdc.gov/foodnet/). The National Molecular Subtyping Network for Foodborne Disease Surveillance (PulseNet)[35] performs DNA "fingerprinting" on bacteria that may be foodborne at http://www.cdc.gov/ncidod/dbmd/pulsenet/pulsenet.htm.

EPIDEMIOLOGIC INVESTIGATION

Forensic epidemiology is a new discipline. However, epidemiologic evidence has proven to be powerful in litigation involving such diverse hazards as swine flu vaccine, Agent Orange, silicone breast implants, and tampons.[36] Several studies describe the use of epidemiological evidence in legal proceedings.[37–43]

Forensic epidemiology was key in documenting the true nature of the Sverdlovsk incident of 1979.[20,44,45] Ninety-six people became ill between April 4 and May 18, 1979, and at least 66 people died of anthrax, in the Russian city of Sverdlovsk (now known as Yekaterinburg) in the eastern foothills of the Urals. Official Soviet reports claimed that the victims had become infected by eating contaminated meat. Autopsy records were seized by government security forces, except for a few handwritten notes and secretly hidden tissue samples, and the dead bodies were buried together in lime, under guard. Epidemiological investigations (questionnaires and disease mapping) conducted 13 years later were able to document that these were cases of inhalation-induced anthrax, and found the point source of the outbreak to be an aerosol emission from Compound 19, a secret military facility.

COMPOSITION OF AN OUTBREAK INVESTIGATION TEAM

The investigation team should have broad scientific expertise to encompass the pathogen(s) likely causing the outbreak, and should include appropriate law enforcement officials. Zoonotic disease investigations should include veterinary and medical epidemiologists, physicians, veterinary clinicians, microbiologists or virologists, and entomologists if the disease is vector-borne. It is important to have personnel highly skilled in interviewing techniques: in prior investigations, personal interviews provided the key information. In domestic outbreaks, local and state health officials can also lend valuable expertise on baseline disease rates and local conditions.

There have been suggestions that organizations such as WHO and OIE should play a role in investigating suspicious incidents internationally because their staff have experience in many countries of interest. However, some scientists are concerned that an inspection or investigatory role for WHO or OIE may inhibit voluntary access to countries and thereby compromise the primary function of these organizations.

EPIDEMIOLOGIC ANALYSIS

Successful epidemiological studies require sound design, sensible analysis, and careful execution. Epidemiologists ensure the quality of their fieldwork in many ways. They seldom collect data without written, standard instruments designed to increase comparability between the study groups. Typical studies use a variety of instruments: abstracts from medical records, personal interview questionnaires, self-administered questionnaires, telephone interview questionnaires, physical examinations, biological specimen collections, and environmental samples.[46]

The first step in any outbreak investigation is to evaluate clinical and laboratory findings to confirm that a disease outbreak has occurred. The next step is to construct a case definition so that the number of cases and the attack rate can be determined. A case is the person or animal in the population or group identified as having the particular disease, or condition under investigation.[6] A variety of criteria may be used to identify cases. Examples of clinical criteria are physician diagnoses, abstracts of clinical records, and a summary of similar clinical findings. Epidemiologic criteria might be exposure history—such as exposure to other cases of a disease or animal vectors—or consumption of a suspected food vehicle. A laboratory criterion would be the accepted diagnostic method for identification of the known etiologic agent. Suspected cases will meet either clinical or epidemiologic criteria, and a definite case usually requires laboratory confirmation.

Accurate development of a case definition is especially important in a potential bioterrorism event, since there may well be an increased potential for psychological consequences with acute psychosomatic symptoms. Such signs and symptoms may be confused with infection or intoxication.[47]

Calculation of the attack rate allows investigators to compare the estimated rate of illness with rates during previous time periods to determine whether the rate constitutes a deviation from the norm. The attack rate or case rate is a cumulative incidence rate (the proportion of a group of people who experience the onset of disease) used for a particular group observed for limited periods under special circumstances such as an epidemic.[6]

ANALYSIS OF EPIDEMIOLOGICAL DATA WITH REGARD TO PERSON, ANIMAL, OR CROP AFFECTED

Once the overall attack rate is computed, it is important to determine whether there is any group more affected by the outbreak, for example, children compared to adults or men compared to women. In the 1997 Taiwan foot-and-mouth disease outbreak, only pigs were affected and not cattle, sheep, or goats on the same premises, pointing to an unusual strain of foot-and-mouth disease virus that was not virulent for ruminants.

ANALYSIS OF DISEASE PATTERNS OVER TIME

Observations are often organized in the form of an epidemic curve in which the number of cases is plotted against a time axis. This disease pattern is an important factor in differentiating a natural outbreak from an intentional event. A point source epidemic is an example of a temporal disease pattern (Figure 10.1). In this case, susceptible individuals are exposed more or less simultaneously to one source of infection; there is a sudden peak in the disease incidence and quick return to normal. Food source outbreaks tend to be point source events. A propagating outbreak, in contrast, is typical of contagious infections like foot-and-mouth disease or pneumonic plague, in which cases continue to occur and each new case may infect others to continue the epidemic (Figure 10.2).

In most naturally occurring outbreaks, other than those that are foodborne, numbers of cases gradually increase as progressively larger numbers of people, animals, or plants come in contact with other patients, animals, fomites (inanimate objects capable of carrying infection from one infected animal to another), and vectors that can spread disease.[19] Epidemic curves may also indicate the time of exposure.

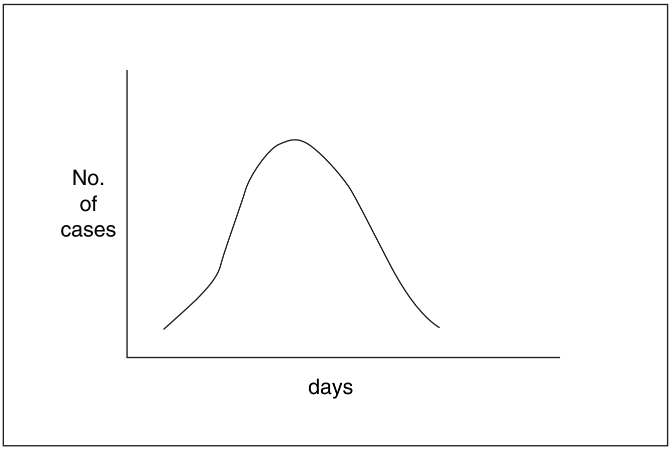


FIGURE 10.1 Example of a temporal pattern of disease typical of a point source epidemic such as a foodborne infection. Susceptible individuals are exposed more or less simultaneously to one source of infection: there is a sudden peak in the incidence of disease and a quick return to normal.

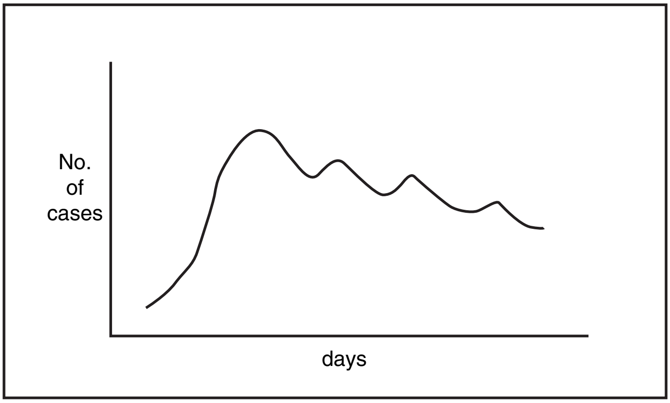


FIGURE 10.2 Example of a propagating outbreak involving a contagious disease such as foot-and-mouth disease virus infection in which new cases continue to occur and each new case may infect others to continue the epidemic.

ANALYSIS OF DISEASE DATA WITH REGARD TO GEOGRAPHICAL DISTRIBUTION

Disease mapping, which provides a visual representation of the geographical distribution of a disease within a population, can often be a valuable tool in determining the source of an outbreak. The most famous example is John

Snow's studies of the cholera epidemic in the Golden Square area of London in 1854.[48] The epidemiology and etiology of cholera were unknown, but through meticulous attention to clinical history and by plotting where his patients lived, Snow demonstrated that cholera was spread through a contaminated water supply. His findings preceded isolation of the cholera bacterium by 44 years. His dot map clearly demonstrated a distinct cluster of cases around the water pump on Broad Street. It was later found that the pump piped water from the River Thames that was contaminated with sewage.[49]

The power of disease mapping was also demonstrated in the 1979 Sverdlovsk anthrax outbreak: a map showed that the majority of victims lived or worked along a narrow geographic strip that extended four kilometers from a military microbiological facility within the city to the city's southern border. When researchers compared maps of the victims' locations during the day to those in the evening hours, they were able to document that many of the scattered plots of infected victims during the day collapsed into a narrow zone in the evening.[44] Even though the Soviet government had claimed that the epidemic was caused by consumption of contaminated meat, this statistically significant mapping pattern could not be explained by a foodborne illness; it strongly implicated a windborne aerosol of anthrax spores at night. In addition, the linear pattern of the disease outbreak and study of meteorological patterns pointed to the military microbiological facility.

In mapping studies, using raw numbers of cases rather than case rates can lead to erroneous conclusions. Cases may be concentrated in large business complexes, factories, or apartment buildings simply due to the large number of people working or living in the area. Approximate population densities and locations can be obtained from census information or from satellite photographs in uncooperative environments such as Sverdlovsk. Host and environmental factors associated with geography that can play a role in initiation of disease include the presence of vectors, climate, population density, nutritional and dietary practices, occupation, recreational habits, economic development, and socially disruptive events such as war and natural disasters.[46]

The emerging field of bioinformatics that integrates biology with multiple databases from other disciplines—such as geography, health, climate, mammalogy, agronomy, and economics—is rapidly changing our ideas of what is possible in understanding complex field situations. As Baker and colleagues have aptly said, "Bioinformatics is a systems approach that allows for retrospective analysis of past events, real time analysis of complex contemporary data sets, and, ideally, positioning for an unpredictable future."[50] Such an approach has already been applied successfully in unraveling the natural history of a newly emerging disease—hantavirus pulmonary syndrome—that illustrates all the difficulties and complexities of responding to a disease incident involving an unknown pathogen that has many of the characteristics of an unusual outbreak as described earlier.[51]

GLOBAL INFORMATION SYSTEMS TOOLS TO ASSIST DISEASE MAPPING

Many commercial software packages, referred to as Geographical Information Systems (GISs), allow display and manipulation of geo-referenced data[52–54] useful in outbreak investigation. Most mapping programs allow construction of multiple map layers with different information in each layer. Two common software packages are MapInfo at http://www.mapinfo.com and ArcView at http://www.esri.com/software/arcview. Other programs available as shareware (at the time of writing) include BUGS and BEAM at http://www.mrc-bsu.cam.ac.uk, SATSCAN (at martink@cortex.uchc.edu), and DISMAPWIN at http://ftp.ukbf.fu-berlin.de/sozmed/DismapWin.html.

Comparison of the disease outbreak location with historical knowledge of regional endemic and epidemic diseases provided through the WHO, Promed, FAO, or OIE global mapping will provide clues as to whether or not the occurrence is a natural event.

CONDUCTING INTERVIEWS

Interviews are often the most important and extensive source of information for determining the source of an outbreak. An epidemiological field investigation should include interviews with outbreak survivors, families of outbreak victims, local physicians, and public health officials involved in the epidemic and in heath care measures to contain the outbreak. The team must acquire the names and addresses of all outbreak victims and obtain permission to interview their families. As the investigation continues, investigators may need to interview the same sources multiple times to obtain accurate and revealing information. Therefore, it is critical that investigation guidelines permit multiple interviews.

Structured interviewing techniques (refined questionnaires) should be used to expedite the process and to obtain consistent results. If the questionnaire requires the respondent to recall distant events, incorporation of memory aids can be invaluable. Personal interviews can incorporate pictures, three-dimensional models, maps, calendars, or timelines to assist respondents with accurate reporting. Personal interviews are labor-intensive and expensive. Telephone interviews typically yield shorter answers, and respondents may tend to favor the first answer when a list of possible answers is read to them, but other differences between telephone and personal interviews are, surprisingly, modest[55]. In a biological terrorism event, cooperation in a telephone interview may be complicated by fear, as respondents tend to be more suspicious of the interviewer or the legitimacy of the study when face-to-face contact is lacking.

EVALUATION OF CLINICAL FINDINGS

The pathogens or toxins that cause anthrax, plague, tularemia, Brucellosis, viral encephalitis, smallpox, botulism, and Staphylococcal enterotoxemia have been weaponized or considered for use in biological weapons programs. Although a certain number of cases of these diseases (except smallpox) occur naturally in the U.S. each year, they still should be evaluated carefully with a certain index of suspicion. Some of these diseases display distinct clinical appearances and pathology depending upon the route of exposure. Analysis of this information may be valuable in suggesting a source for the outbreak and whether or not foul play may be involved.

An excellent review of the diagnosis and management of patients exposed to biological warfare agents was recently published; the reader is referred to that publication for additional information[56]. The present discussion will focus on specific aspects of clinical disease that point to intentional misuse.

Anthrax has three clinical presentations: cutaneous, gastrointestinal, and inhalational. Inhalational anthrax begins with fever, malaise, nonproductive cough, and chest discomfort and progresses to severe respiratory distress, bacteremia, septic shock, metastatic infection, and death. Physical findings are nonspecific. Chest radiographs typically show a widened mediastinum[57]. Inhalational anthrax, as found in Sverdlovsk and the October 2001 postal attacks in the U.S., is highly indicative of deliberate exposure. Although gastrointestinal anthrax occurs in countries lacking well-developed meat inspection systems, in the U.S. it is rare and would be considered suspicious.

There are three clinical presentations of plague–bubonic, primary septicemic, and pneumonic. Patients with pneumonia due to *Yersinia pestis* present with high fever, headache, myalgia, and productive cough with bloody sputum, which progresses rapidly to sepsis[58]. Plague has been endemic in the western U.S. since 1900, with 5–15 cases of human disease occurring each year; however, only 2% of plague has been pneumonic. Therefore pneumonic plague in the U.S. would raise the suspicion of an intentional event.

Tularemia presents as ulcero-glandular disease when the initiating event is penetration through the skin or mucous membranes. Typhoidal and septicemic disease, which presents with fever and weight loss, usually develops after inhalation of infectious aerosols. Pneumonia occurs with either form, but more likely follows inhalation[59]. Therefore, a tularemia outbreak with a high incidence of pneumonic disease should be investigated with intentional misuse in mind.

Patients with inhalation-induced Staphylococcal B enterotoxemia have difficult breathing and chest pain: in severe cases there may be pulmonary edema progressing to adult respiratory distress syndrome.[56] While some cases of inhalation-induced disease have gastrointestinal symptoms, naturally occurring cases of staphylococcal food poisoning cases do not present with pulmonary signs or symptoms.

Ricin induces severe pulmonary disease after inhalation exposure[60] as contrasted with severe gastrointestinal signs of nausea, vomiting, diarrhea, and rectal hemorrhage[56] after oral administration. The signs and symptoms of Botulinum intoxication and *Brucella* infection do not differ with route of administration.[56]

When traditional pathogens are deliberately used in a typical route of exposure, for example when Rajneesh cult members contaminated a salad bar with *Salmonella*, the clinical syndrome induced by misuse may present an identical clinical picture to naturally occurring disease.[16]

INVESTIGATION OF ANIMAL DISEASE OUTBREAKS

Epidemiological investigations of animal disease outbreaks will have many similarities to those in humans. However, interviewing may be of even more significance than in outbreaks of human illness. One-on-one interviews with farmers are often invaluable in pinpointing the source of a problem. Farmers are usually well aware of the costs incurred from the occurrence and spread of an exotic disease. They often have local knowledge of farmers who buy animals of questionable provenance, those who smuggle livestock across borders, and those with poor management practices, such as feeding uncooked food garbage (the suggested cause of the 2001 foot-and-mouth disease epidemic in Great Britain), which could contribute to initiation or spread of disease. The associations of tuberculosis-positive cattle with the Texas-Mexico border, Newcastle disease with smuggled parrots, and avian influenza with live bird markets are well known.[61]

Movement of people and contaminated fomites must not be overlooked in disease outbreak investigation. Farming is a social enterprise: neighbors milk each other's cattle, help each other put up hay or silage, and spread manure which may be contaminated on another's field. Food, pasture, water, and bedding must be checked. Movements of people, traffic, and animals should be evaluated. Animals and facilities of neighboring farms should also be inspected. Weather and wind patterns can also be a significant factor in the spread of diseases such as foot-and-mouth that are often aerosol-induced.

In diseases of domestic livestock, less is known about different clinical presentations due to agent route of exposure. Etiologic strain comparisons to known endemic mapping may provide better clues as to uniqueness of an outbreak.

One of the challenges in investigations of animal disease outbreaks is tracking exposed animals and animal products. Individual animal identification, such as the cattle passports used in the European Union and barcode labeling to preserve the identity of products, are invaluable in tracking disease outbreaks. Availability of GIS information linked to farms, livestock data, livestock markets, and feed suppliers proved helpful in the foot-and-mouth outbreak in Great Britain in 2001.

INVESTIGATION OF CROP DISEASE OUTBREAKS

As with other outbreaks, it is necessary to have a clear understanding of what naturally occurring disease looks like in order to appreciate the unusual. An excellent review of natural plant diseases was given by Rogers;[62] for convenience, the major points of that paper are summarized here. Just as in human or animal illness, crop diseases have typical epidemic curves depending on the interaction of the pathogen and its host. Plant pathogens that undergo one lifecycle in one crop growing season are called monocyclic. If there are several pathogen lifecycles in a growing season, a polycyclic epidemic curve is expected.

The intensity of a crop disease epidemic depends on several factors: the virulence of the pathogen, the environment, and the susceptibility of the host plant. Pathogens that infect plants through their leaves and produce spores that are blown by the wind, and pathogens that are spread by flying insects are most likely to trigger rapid epidemics. Development of fungal diseases is greatly affected by weather, being impacted by even small changes in temperature and humidity. Genetic uniformity of a crop predisposes to rapid spread of a pathogen; heterogeneous populations rarely experience epidemic disease.

Sources of naturally occurring plant infections include residual spores surviving in the soil from previous growing seasons, over-wintering of the pathogen on another crop, diseased crop residues that were not properly disposed of, and spread from wild plants.

Most anti-crop biological warfare programs have focused on airborne fungal pathogens of major food crops that are prone to rapidly developing epidemics within a single growing season.[62,63] Examples include cereal rusts and smuts, potato late blight, and brown spot of rice. Most soilborne diseases of crops, whether due to viruses, Mycoplasmas, or fungi, are typically slow to spread and would be less effective in a biological attack.

The epidemiology of the major diseases of significant food crops is well enough known to accurately predict the likely development of endemic disease problems within a growing season after taking into account climactic conditions, pathogen reservoirs, and plant types. Recent technological advances such as satellite imagery and mathematical modeling can assist in such analyses.

SPECIMEN COLLECTION

Investigation should be initiated as soon as the outbreak is identified. Rapid identification of the causative agent and the likely source or mode of transmission is essential. Specimens obtained in the acute phase of disease prior to initiation of therapy are most likely to yield the viable pathogen. The chance of virus recovery is best during the first three days of onset and is greatly reduced beyond five days. Decontamination procedures, taken to prevent the spread of disease, may destroy environmental samples if sampling is not initiated quickly. Autopsy samples should be collected as soon after death as possible. A multidisciplinary team that includes chain of custody specialists can provide optimal sampling and handling techniques. When investigating suspect cases, it is important to keep an open mind as to sources and etiology to ensure that adequate clinical and environmental samples are taken to eliminate uncertainty.

Investigators must bear in mind that the primary focus in a disease outbreak must be to get the disease under control and to limit casualties. Advanced coordination and planning of law enforcement agencies with public health, animal health, and crop protection agencies is essential to ensure that outbreak control and criminal investigation can proceed simultaneously and effectively.

ENVIRONMENTAL SAMPLING

Adequate sampling of environmental fomites and rodents can resolve suspected intentional events. Hantavirus infection of Native Americans in the Southwest[64] and the tularemia outbreak in Kosovo[65] are two examples in which environmental sampling, combined with epidemiological and clinical evaluations, attributed suspicious outbreaks to natural causes.

SAMPLE HANDLING

Successful laboratory confirmation requires collection of appropriate and adequate specimens, proper packaging, and rapid transport to the appropriate laboratory, while ensuring the chain of custody of the specimens and following appropriate biological safety procedures. It is advisable to consult with the receiving laboratories about all aspects of sample collection prior to collection of the samples if possible.

Biological safety precautions need to be observed while conducting the investigation, both to protect the investigator and to prevent unintentional spread of disease. In suspicious outbreaks, appropriate triage of samples may have to be performed to exclude radiological and toxic chemical exposure of the investigator.

Each sample collected should be assigned a unique identification number that is linked with a laboratory request form. If the sample is a clinical sample, the patient's name should be present on all specimens, epidemiological data forms, and the laboratory request form. Appropriate sealing of laboratory samples and/or transport media can maintain the security and integrity of samples during shipment, identify samples whose identity or integrity may have been compromised, and identify and link samples with the information required for accurate analysis and reporting of test results.

SPECIMEN STORAGE

Biological specimens must be handled properly to preserve bacterial, viral, or toxin viability and integrity. In an outbreak investigation, it should be mandatory to consult with the receiving laboratory about proper handling of the expected specimens before beginning the field investigation. Recommended conditions should be maintained throughout transport to the laboratory. Bacterial specimens should be placed into appropriate transport media at recommended temperatures to ensure pathogen viability while minimizing overgrowth of other background flora. If incubation must be delayed, most bacteria can be stored at 4–6°C for short periods of time. Some bacteria, such as *Shigella* species, are cold-sensitive and may be adversely impacted by refrigeration. Viral specimens may be maintained in viral-specific media for several days at 4–6°C: some viral pathogens are destroyed by freezing.

LABORATORY ANALYSIS

Laboratory analysis of samples from a suspicious outbreak will require strict handling and analytical procedures. The analyst must have documented and appropriate training and experience with that type of sample. Analyses will require calibrated equipment with standardized reagents. Internal standards, replicate analysis, and blind testing should be used for the highest possible credibility of results.[66–70]

It is important that the receiving laboratory maintain and document the chain of custody of the sample from the time of receipt until the report is finalized and clearance has been obtained from the submitting investigator. If information gained during the investigation suggests that the outbreak is due to a criminal event, the remaining sample should be retained in a secure environment to enable its use as evidence. The laboratory should record the date and time when the sample is received, the name of the person receiving the sample, and a record of the condition of the sample.

SEROLOGY

This important area has been covered in the chapter by Steven Schutzer elsewhere in this volume.

EVALUATION OF THE ETIOLOGIC AGENT

The type of pathogen or toxin detected in an outbreak can signal intentional misuse. For example, Staphylococcal enterotoxin B (SEB) has been weaponized due to its relative stability. But Staphylococcal enterotoxin A is more commonly found as a naturally occurring cause of foodborne illness. Therefore, the occurrence of illness due to SEB should raise considerable interest in an outbreak investigation.

Presence of multiple pathogens within a clinical specimen or within one outbreak of illness can also indicate unnatural occurrence. Analysis of preserved human tissue specimens from Sverdlovsk contained four different strains of the anthrax bacillus.[20] Such multistrain occurrence is a highly improbable event in a naturally occurring disease outbreak, and was considered strong evidence that the anthrax was manufactured.

Microbes are dynamic organisms, and are therefore constantly changing in response to natural selection pressures in their environment. It is often possible to determine the laboratory or geographic origin and character of microbes by use of strain typing, molecular biology tools, and antibiotic resistance patterns. For example, there are seven major serotypes of foot-and-mouth disease virus, and these have definite geographic distributions throughout the world. Ribotyping of *Yersinia pestis* was useful in ruling out a terrorist attack by Kashmir separatists in the pneumonic plague epidemic in Surat in 1994. Similarly, when parties to the war in Northern Bosnia accused each other of biological warfare during the 1995 tularemia epidemic, strain typing revealed the etiologic agent to be type B *Francisella tularensis* var *palaeartica*, which is endemic in Bosnia.[65] Pathogen banks at various diagnostic reference centers and collaborating centers associated with the OIE and WHO maintain strain typing and molecular characterization information.

If the etiologic agent differs from agents found naturally, detailed molecular analysis may be able to indicate whether or not the organism has been genetically or otherwise modified.

ANALYSIS OF METEOROLOGICAL AND CLIMATIC CONDITIONS

After geographical analysis in Sverdlovsk showed the cases to be tightly clustered in a linear plume southeast of the city, meteorological analysis pinpointed the date of exposure to 2 April 1979.[44] In addition, wind direction, rainfall, and temperature history are all useful clues as to whether or not a disease would likely occur naturally. Climate influences infectious disease patterns because pathogens and their vectors are sensitive to temperature, rainfall, and other ambient conditions. This is shown by the characteristic geographic distribution and seasonal variation of many infectious diseases. Mosquito-vectored diseases, such as Dengue, malaria, and West Nile fever are seen primarily during warm periods; influenza increases in cool weather; meningitis is associated with dry environments; and cryptosporidiosis is associated with heavy rainfall.[71] Wind direction and speed also influence the pattern of disease spread, as noted at Sverdlovsk and in foot-and-mouth disease outbreaks.

ECONOMIC ANALYSIS

The primary impacts of the crop or livestock pathogens likely to be used as weapons are economic. These losses have two components: those associated with outbreak control, and those triggered by international trade embargos of animals or plants and their products. Confirmation of even one case of foot-and-mouth disease in a country normally free of this infection immediately stops all export of cloven-hoofed animals, meat, and animal byproducts, and also impacts seemingly nonrelated exports, such as used farm equipment and antiques.

Current world production, market, and trade reports for major commodities are available by commodity listing and by country of origin through the USDA Foreign Agricultural Service website http://www.fas.usda.gov/currwmt.html. Reports are available for livestock and poultry, cotton, dairy, fishery products, forest products, grain, sugar, tobacco, and other commodities. The world market and trade reports provide the latest analysis and data on a number of agricultural commodities, outlining the current supply, demand, and trade estimates for both the U.S. and many major foreign countries. Analyses of specific disease occurrences are also available on this website. Indications of economic benefit to a particular party after a disease outbreak may occasionally be useful in raising the index of suspicion about potential sources of infection.

MEDIA REPORTS

Disease reports in news media often occur when countries are denying disease outbreaks for trade and economic purposes. Before and during the 1997 foot-and-mouth disease outbreak in Taiwan, the People's Republic of China did not report the presence of foot-and-mouth disease to OIE or FAO, but newspaper reports described the presence of disease, which was a possible source for the Taiwan outbreak.

However, media reports are not always accurate and should be cross-checked and verified with other sources of information. PATHFINDER is a Web and news service search program developed by APHIS that captures and stores animal health data for subsequent analysis.

AVAILABLE DOCUMENTS

Lists of victims, hospital admissions, medical records, veterinary records, meteorological data from local airports and weather stations, laboratory accession data, import permits, and sale barn receipts may all provide important epidemiologic information on an outbreak. Information on facilities from which an accidental release could have taken place may be obtained from records of arms control declarations or satellite imagery.

EVALUATION OF SCIENTIFIC LITERATURE

Terrorist or biological warfare attacks often do not take place in isolation. Successful attacks usually follow successful research and field trials. "Any technology has a parenthood and genealogy attached to it."[61] Similar research may be conducted in different facilities by students or teachers of the perpetrators. There are exceptions, however, such as the New Zealand farmers who infected pet rabbits with rabbit calicivirus (which had been smuggled into the country in dried blood on a handkerchief) to amplify the virus, and then fed ground livers from these animals to wild rabbits to trigger a highly lethal epidemic. Another example is the Rajneesh attack in The Dalles, Oregon, for which there were sufficient precedents in natural Salmonella outbreaks to render research unnecessary.

TRAINING OR WORK EXPERIENCE

A review of biocrimes[18] demonstrated a disproportionate number of physicians, nurses, technical laboratory workers, biomedical students, and students of microbiology associated with such events. In these cases, education provides capability and access to biological materials, and equipment adds capacity.

EQUIPMENT OR POTENTIAL MEANS OF DELIVERY

Production facilities scaled to produce the quantity and quality of etiologic agent typical of military biological warfare commonly operate at the P3 or P4 biological containment level and have specialized production equipment, such as 20-liter or larger fermenters, often with continuous flow systems; centrifugal separators with a flow rate of 100 liters or more per minute; cross-(tangential) flow filtration equipment with a total filtration area equal to or greater than 1 square meter; and freeze-drying equipment with a condenser capacity of 10 kg of ice or greater in 24 hours. Other indicators of work with high-consequence pathogens are personal protection full or half-suits, class III biological safety cabinets or isolators with similar performance standards, and chambers designed for aerosol challenge. Equipment of these types is examined in UN weapons inspections. The U.S. limits their export to questionable countries.

However, many biological crimes and terrorist attacks have been conducted with a much smaller footprint than that of a state-sponsored biological weapons program. The Rajneesh attack on a salad bar, which sickened more than 600 people,[16] and the Shigella contamination of muffins[17] required only a small clinical incubator. The New Zealand rabbit calicivirus epidemic that began with a bloodstained handkerchief, and started an epidemic that killed millions of wild rabbits required only a kitchen blender and one infected liver. Lack of conventional production capabilities should not rule out a possible intentional event.

SUMMARY

Most outbreaks of high-consequence pathogens and toxins are likely to have a normal explanation when all the data are evaluated. In all likelihood, no one piece of data or analysis will prove or disprove the existence of an intentional event. That proof will require multiple supporting pieces of evidence. The implications of a proven attack or biocrime are so far reaching that any investigation resulting in such a conclusion must be so thorough as to survive the most rigorous of examinations.[61]

REFERENCES

1. United Nations. Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction. April, 1972.
2. United Nations. Fourth Review Conference of the Parties to the Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction, Final Declaration, Final Document, BWC/CONF.IV/9, 6 December 1996.
3. United States. Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA Patriot Act) Act of 2001. Section 817. Expansion of the Biological Weapons Statute. 2001.
4. United Nations. The Second Review Conference of the States Parties to the Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction, Final Document, Geneva, 8–26 September 1986, BWC/CONF.II/13. 1996.
5. United Nations. Ad Hoc Meeting of Scientific and Technical Experts from States Parties to the Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction, BWC/CONF.II/EX/2, 21 April 1987.
6. Last JM, Abramson JH. International Epidemiological Association. *A Dictionary of Epidemiology.* New York: Oxford University Press; 1995.
7. World Health Organization. *Laboratory Biosafety Manual.* Geneva: World Health Organization; 2003.
8. Geissler E, Woodhall JP. Reporting of Outbreaks of Disease Under BTWC. Confidence Building Measures. *In* Dando M, Pearson GS, Kriz B (eds). *Scientific and Technical Means of Distinguishing Between Natural and Other Outbreaks of Disease. Proceedings of the NATO Advanced Research Workshop*, Prague, Czech Republic, 18–20 October 1998. 2001.
9. Pearson GS. The Importance of Distinguishing Between Natural and Other Outbreaks of Disease. *In* Dando M, Pearson GS, Kriz B (eds). *Scientific and Technical Means of Distinguishing Between Natural and Other Outbreaks of Disease. Proceedings of the NATO Advanced Research Workshop*, Prague, Czech Republic, 18–20 October 1998. 2001.
10. World Health Organization. Public Health Response to Biological and Chemical Weapons: WHO guidance. Geneva: 2004.
11. Australian Group. List of Plant Pathogens for Export Control http://www.australiagroup.net/en/control_list/plants.htm. Web. 2004.

12. Australian Group. List of Animal Pathogens for Export Control http://www.australiagroup.net/en/control_list/animal.htm. Web. 2003.
13. Australian Group. List of Biological Agents for Export Control http://www.australiagroup.net/en/control_list/bio_agents.htm. Web. 2004.
14. Select Agent Act, title 42, Code of Federal Regulations part 73. 2003.
15. Department of Health and Human Services. Possession, use and transfer of select agents and toxins. 42 CFR Part 73. 2003.
16. Torok TJ, Tauxe RV, Wise RP, Livengood JR, Sokolow R, Mauvais S, Birkness KA, Skeels MR, Horan JM, Foster LR. A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. *JAMA* 1997;278:389–395.
17. Kolavic SA, Kimura A, Simons SL, Slutsker L, Barth S, Haley CE. An outbreak of Shigella dysenteriae type 2 among laboratory workers due to intentional food contamination. *JAMA* 1997;278:396–398.
18. Carus WS. Bioterrorism and Biocrimes: The Illicit Use of Biological Agents in the 20th Century Center for Counterproliferation Research, National Defense University, Washington, DC, July 1999.
19. Pavlin JA. Epidemiology of bioterrorism. *Emerg Infect Dis* 1999;5:528–530.
20. Jackson PJ, Hugh-Jones ME, Adair DM, Green G, Hill KK, Kuske CR, Grinberg LM, Abramova FA, Keim P. PCR analysis of tissue samples from the 1979 Sverdlovsk anthrax victims: The presence of multiple Bacillus anthracis strains in different victims. *Proc Natl Acad Sci* U S A 1998;95:1224–1229.
21. Dando M, Whitby S. Distinguishing Between Natural Disease Outbreaks and Deliberate Attacks on Humans Using Biological Weapons. *In* Dando M, Pearson GS, Kriz B (eds). *Scientific and Technical Means of Distinguishing Between Natural and Other Outbreaks of Disease. Proceedings of the NATO Advanced Research Workshop*, Prague, Czech Republic, 18–20 October 1998, 2001.
22. O.I.E. Terrestrial Animal Health Code 2003. Chapter 1.3.6.1. Surveillance and Monitoring of Animal Health. 2004.
23. Preslar DB. The Role of Disease Surveillance in the Watch for Agro-terrorism or Economic Sabotage, AHEAD/ILIAD, November 2000.
24. Freedman DO, Kozarsky PE, Weld LH, Cetron MS. GeoSentinel: The global emerging infections sentinel network of the International Society of Travel Medicine. *J Travel Med* 1999; 6:94–98.
25. Leder K, Sundararajan V, Weld L, Pandey P, Brown G, Torresi J. Respiratory tract infections in travelers: A review of the GeoSentinel surveillance network. *Clin Infect Dis* 2003;36: 399–406.
26. Koo D, Wetterhall SF. History and current status of the National Notifiable Diseases Surveillance System. *J Public Health Manag Pract* 1996;2:4–10.
27. Summary of notifiable diseases—United States, 2001. *MMWR Morb Mortal Wkly Rep* 2003; 50:i–108.
28. Bean NH, Martin SM, Bradford H, Jr. PHLIS: An electronic system for reporting public health data from remote sites. *Am J Public Health* 1992;82:1273–1276.
29. Hajjeh RA, Relman D, Cieslak PR, Sofair AN, Passaro D, Flood J, Johnson J, Hacker JK, Shieh WJ, Hendry RM, Nikkari S, Ladd-Wilson S, Hadler J, Rainbow J, Tappero JW, Woods CW, Conn L, Reagan S, Zaki S, Perkins BA. Surveillance for unexplained deaths and critical illnesses due to possibly infectious causes, United States, 1995–1998. *Emerg Infect Dis* 2002;8:145–153.
30. Briese T, Jia XY, Huang C, Grady LJ, Lipkin WI. Identification of a Kunjin/West Nile-like flavivirus in brains of patients with New York encephalitis. *Lancet* 1999;354:1261–1262.
31. United States General Accounting Office. West Nile Virus Outbreak: Lessons for Public Health Preparedness. United States General Accounting Office, Report to Congressional

Requesters. September 2000. GAO/HEHS-00–180. http://www.gao.gov/new.items/he00180.pdf. 2000.

32. Guarner J, Johnson BJ, Paddock CD, Shieh WJ, Goldsmith CS, Reynolds MG, Damon IK, Regnery RL, Zaki SR. Monkeypox transmission and pathogenesis in prairie dogs. *Emerg Infect Dis* 2004;10:426–431.
33. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food—selected sites, United States, 2003. *MMWR Morb Mortal Wkly Rep* 2004;53:338–343.
34. Yang S. FoodNet and Enter-net: Emerging surveillance programs for foodborne diseases. *Emerg Infect Dis* 1998;4:457–458.
35. Ransom G, Kaplan B. USDA uses PulseNet for food safety. *J Am Vet Med Assoc* 1998;213:1107.
36. Loue S. *Case Studies in Forensic Epidemiology.* New York: Kluwer Academic/Plenum Publishers; 2002.
37. Loue S, NetLibrary I. Forensic epidemiology. *A Comprehensive Guide for Legal and Epidemiology Professionals.* Carbondale: Southern Illinois University Press; 1999.
38. Hertzman PA, Blevins WL, Mayer J, Greenfield B, Ting M, Gleich GJ. Association of the eosinophilia-myalgia syndrome with the ingestion of tryptophan. *N Engl J Med* 1990;322: 869–873.
39. Steere AC, Malawista SE, Snydman DR, Shope RE, Andiman WA, Ross MR, Steele FM. Lyme arthritis: An epidemic of oligoarticular arthritis in children and adults in three Connecticut communities. *Arthritis Rheum* 1977;20:7–17.
40. Benach JL, Bosler EM, Hanrahan JP, Coleman JL, Habicht GS, Bast TF, Cameron DJ, Ziegler JL, Barbour AG, Burgdorfer W, Edelman R, Kaslow RA. Spirochetes isolated from the blood of two patients with Lyme disease. *N Engl J Med* 1983;308:740–742.
41. Reingold AL, Kane MA, Murphy BL, Checko P, Francis DP, Maynard JE. Transmission of hepatitis B by an oral surgeon. *J Infect Dis* 1982;145:262–268.
42. Mahoney FJ, Hoge CW, Farley TA, Barbaree JM, Breiman RF, Benson RF, McFarland LM. Communitywide outbreak of Legionnaires' disease associated with a grocery store mist machine. *J Infect Dis* 1992;165:736–739.
43. Canter J, Mackey K, Good LS, Roberto RR, Chin J, Bond WW, Alter MJ, Horan JM. An outbreak of hepatitis B associated with jet injections in a weight reduction clinic. *Arch Intern Med* 1990;150:1923–1927.
44. Meselson M, Guillemin J, Hugh-Jones M, Langmuir A, Popova I, Shelokov A, Yampolskaya O. The Sverdlovsk anthrax outbreak of 1979. *Science* 1994;266:1202–1208.
45. Walker DH, Yampolska O, Grinberg LM. Death at Sverdlovsk: What have we learned? *Am J Pathol* 1994;144:1135–1141.
46. Gerstman BB. Selected Epidemiologic Disease Concepts. *In* Gerstman BB (ed). *Epidemiology Kept Simple: An Introduction to Classic and Modern Epidemiology.* New York: Wiley; 1998.
47. Holloway HC, Norwood AE, Fullerton CS, Engel CC, Jr., Ursano RJ. The threat of biological weapons. Prophylaxis and mitigation of psychological and social consequences. *JAMA* 1997;278:425–427.
48. Snow J. Snow on Cholera. A reprint of two papers. London: The Commonwealth Fund; 1936.
49. Rothman KJ, Greenland S. Types of Epidemiologic Studies. *In* Rothman KJ, Greenland S (eds). *Modern Epidemiology.* Lippincott-Raven; 1998.
50. Baker RJ, Phillips CJ, Bradley RD, Burns JM, Cooke D, Edson GF, Harragan DR, Jones C, Monk RR, Montford JT, Parker NC. Bioinformatics, Museums and Society: Integrating Biological Data for Knowledge-Based Decisions. Occasional Papers, Museum of Texas Tech Univ. 187:i8. 1998.
51. Yates TL, Mills JN, Parmenter CA, Ksiazek TG, Parmenter RR, Calisher C, Nichol ST, Abbot K, Young J, Morrison M, Beaty M, Dunnum JL, Baker RJ, Peters CJ. The ecology and

evolutionary history of an emergent disease: Hantavirus pulmonary syndrome. *Bioscience* 2002;989–998.

52. Geographical Information Systems (GIS). Mapping for epidemiological surveillance. *Wkly Epidemiol Rec* 1999;74:281–285.
53. Briggs DJ, Elliott P. The use of geographical information systems in studies on environment and health. *World Health Stat Q* 1995;48:85–94.
54. Scholten HJ, de Lepper MJ. The benefits of the application of geographical information systems in public and environmental health. *World Health Stat Q* 1991;44:160–170.
55. Hartge P, Cahill JI. Field Methods in Epidemiology. *In* Rothman KJ, Greenland S (eds). *Modern Epidemiology.* Lippincott-Raven; 1998.
56. Franz DR, Jahrling PB, McClain DJ, Hoover DL, Byrne WR, Pavlin JA, Christopher GW, Cieslak TJ, Friedlander AM, Eitzen EM, Jr. Clinical recognition and management of patients exposed to biological warfare agents. *Clin Lab Med* 2001;21:435–473.
57. Friedlander AM, Welkos SL, Pitt ML, Ezzell JW, Worsham PL, Rose KJ, Ivins BE, Lowe JR, Howe GB, Mikesell P. Postexposure prophylaxis against experimental inhalation anthrax. *J Infect Dis* 1993;167:1239–1243.
58. Butler T. Yersinia Species, Including Plague. *In* Mandell GL, Douglas RG, Bennett JE, Dolin R (eds). *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases.* Philadelphia: Churchill Livingstone; 2000.
59. Mandell GL, Douglas RG, Bennett JE, Dolin R. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases.* Philadelphia: Churchill Livingstone; 2000.
60. Wilhelmsen CL, Pitt ML. Lesions of acute inhaled lethal ricin intoxication in rhesus monkeys. *Vet Pathol* 1996;33:296–302.
61. Hugh-Jones M. Distinguishing Natural and Unnatural Outbreaks of Animal Diseases. *In* Dando M, Pearson GS, Kriz B (eds). *Scientific and Technical Means of Distinguishing Between Natural and Other Outbreaks of Disease. Proceedings of the NATO Advanced Research Workshop*, Prague, Czech Republic, 18–20 October 1998. 2001:63–74.
62. Rogers P, Whitby S, Dando M. Biological warfare against crops. *Sci Am* 1999;280:70–75.
63. Whitby S, Rogers P. Biological warfare-implications of the Iraq and US programmes. *Defense Analysis* 1997;13:303–318.
64. Wells RM, Young J, Williams RJ, Armstrong LR, Busico K, Khan AS, Ksiazek TG, Rollin PE, Zaki SR, Nichol ST, Peters CJ. Hantavirus transmission in the United States. *Emerg Infect Dis* 1997;3:361–365.
65. Sohns TM, Finke EJ. Natural and Other Biological Risks from a Medical Service Perspective. *In* Dando M, Pearson GS, Kriz B (eds). *Scientific and Technical Means of Distinguishing Between Natural and Other Outbreaks of Disease. Proceedings of the NATO Advanced Research Workshop*, Prague, Czech Republic, 18–20 October 1998. 2001:63–74.
66. Wheelis M. Will the "New Biology" lead to new weapons? *Arms Control Today* 2004;34:6–13.
67. Kirsch EW. Daubert v. Merrell Dow Pharmaceuticals: Active judicial scrutiny of scientific evidence. *Food Drug Law J* 1995;50:213–234.
68. Masten J, Strzelczyk JJ. Admissibility of scientific evidence post-Daubert. *Health Phys* 2001;81:678–682.
69. Budowle B, Schutzer SE, Einseln A, Kelley LC, Walsh AC, Smith JA, Marrone BL, Robertson J, Campos J. Public health. Building microbial forensics as a response to bioterrorism. *Science* 2003;301:1852–1853.
70. International Standards Organization (ISO)/International Electrotechnical Commission (IEC), *ISO/IEC 17025, General requirements for the competence of testing and calibration laboratories* (1999) American National Standards Institute, New York, NY. 1999.
71. Committee on Climate EIDaHHBoASaCNRC. *Under the Weather: Climate, Ecosystems, and Infectious Disease.* National Academies Press; 2001.

CHAPTER 11

Forensic Handling of Biological Threat Samples in the Lab

JOHN W. EZZELL
U.S. Army Medical Research Institute of Infections Diseases, Ford Detrick, Maryland

INTRODUCTION

Reviewed within this chapter are issues dealing with handling biological threat samples within the laboratory. This seemingly simple subject becomes paramount when forensic issues arise concerning how samples were collected, handled, transported, stabilized, extracted, analyzed, secured, archived, and disposed of, should data from such manipulations be presented in a court of law.[1,2] It is not the purpose of this chapter to provide a "cookbook" of standard operating procedures (SOPs) to address the step-by-step handling of biological threat samples (Figure 11.1), but rather to touch on issues that will influence a laboratory's implementation of an operational strategy for sample handling that will withstand political, scientific, and legal scrutiny in a court of law. Many of these issues were identified as a result of the U.S. Army Medical Research Institute of Infectious Diseases' (USAMRIID) experience with numerous biological threat samples, the Federal Bureau of Investigation (FBI), prosecuting and defense attorneys, and with quality systems.

USE OF TEST PLANS TO IMPLEMENT OPERATIONAL STRATEGIES

PRINCIPLES

Biological samples that are implicated in or collected as part of an investigation of bioterrorism should be handled and analyzed by carefully laid-out operational strategies that conform to a quality standard.[3–5] This often requires

Microbial Forensics
Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

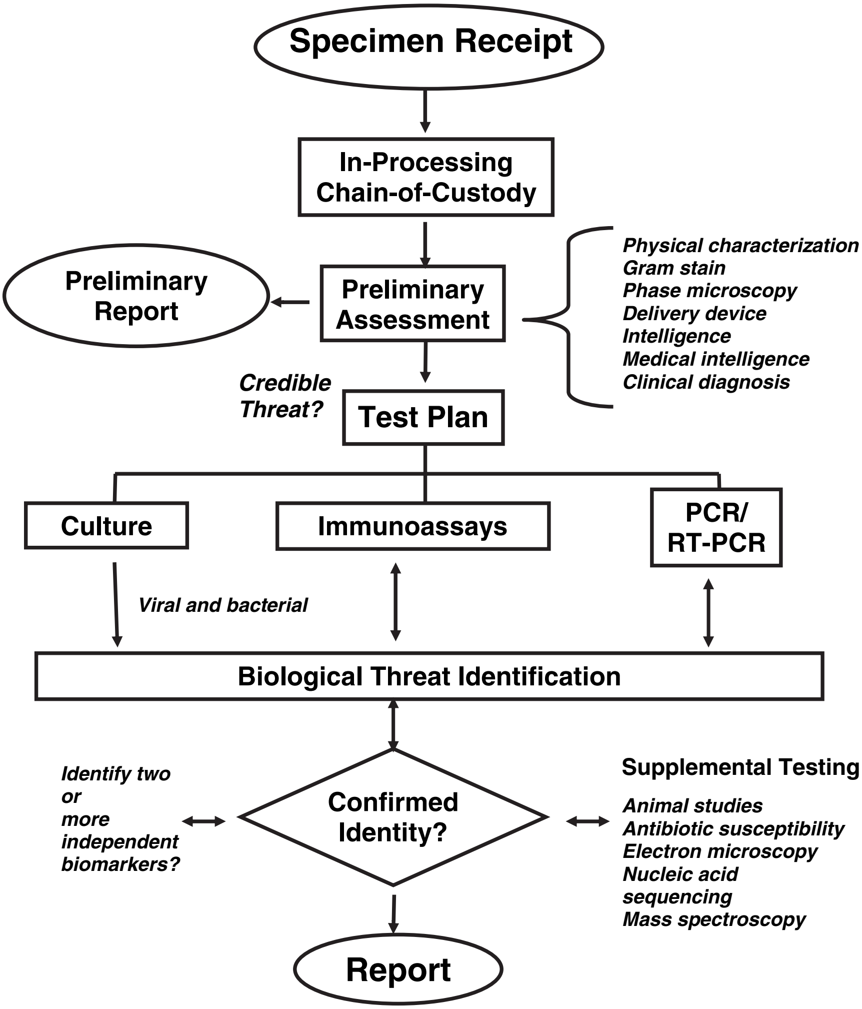

Specimen Receipt
In-Processing Chain-of-Custody
Preliminary Report
Preliminary Assessment
Physical characterization
Gram stain
Phase microscopy
Delivery device
Intelligence
Medical intelligence
Clinical diagnosis
Credible Threat?
Test Plan
Culture
Immunoassays
PCR/ RT-PCR
Viral and bacterial
Biological Threat Identification
Identify two or more independent biomarkers?
Confirmed Identity?
Supplemental Testing
Animal studies
Antibiotic susceptibility
Electron microscopy
Nucleic acid sequencing
Mass spectroscopy
Report

FIGURE 11.1 Flow of specimens in analytical laboratory (from ref. 9, with permission).

preparation of test plans for incoming samples or sample sets. Such test plans help ensure compliance with the laboratory's operational strategy, taking into consideration the samples' unique properties, available quantities, analytical requirements, and confounding circumstances, while protecting forensic evidence. Without a test plan, excessive use of sample material may result in inadequate quantities for subsequent archiving, retesting, testing for other threat agents, or confirmation assays. The test plan should thus reflect calculations

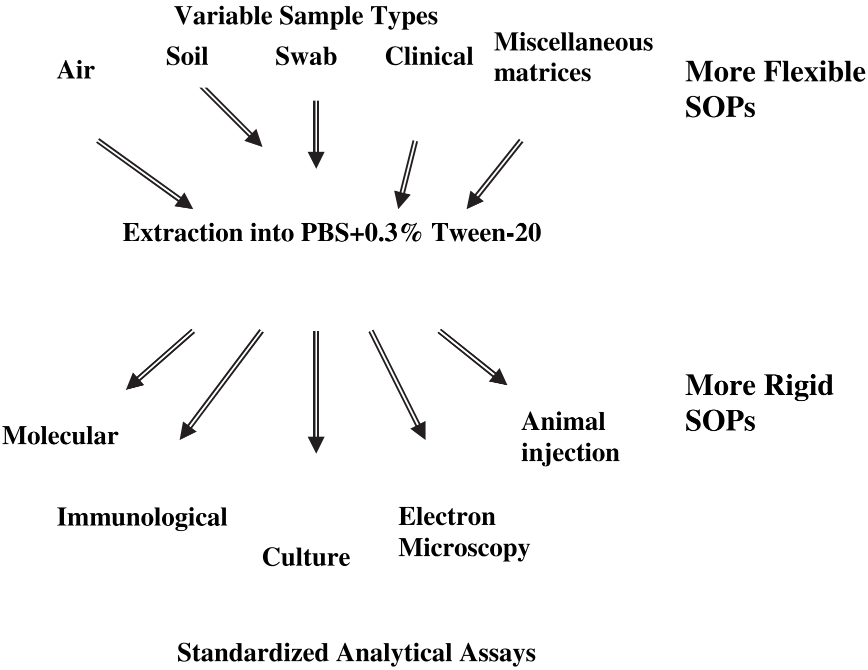

Variable Sample Types
Air
Soil
Swab
Clinical
Miscellaneous matrices
More Flexible SOPs
Extraction into PBS+0.3% Tween-20
Molecular
Immunological
Culture
Electron Microscopy
Animal injection
More Rigid SOPs
Standardized Analytical Assays

FIGURE 11.2 Analytical strategy using flexible and rigid SOPs.

to determine quantities needed to best meet test requirements and to determine whether the operational strategy should be modified due to limited quantities.

Development and implementation of a test plan encompass three principles: 1) State what actions will be performed, 2) perform the stated actions, and 3) document the actions as performed along with any deviations from the test plan. In laboratories where analyses of certain sample types have become routine, test plans may merely follow a series of well-established or validated SOPs. However, attempts to implement an operational strategy with such stringent SOPs can be a daunting task for the vast myriad of biological threat agents, types of matrices, assay inhibitors, stabilization requirements, variations in concentration, etc. Therefore, flexibility may be required in the initial phases of a test plan so as to allow modification of the operational strategy and SOPs to accommodate variations or deviations (Figure 11.2). For example, during the investigations after the *Bacillus anthracis* spore-contaminated mail incident in the fall of 2001, items were received for analysis including clothing, computer keyboards, various forms of garbage, etc., for which there were no validated SOPs. In order to accommodate such variations in sample matrix type, a strategy was adopted by a number of laboratories in which samples were processed so as to convert all sample forms into a common sample matrix (i.e., suspension or extraction in phosphate-buffered saline + 0.3% Tween 20). This was done by using more flexible SOPs; after samples were extracted into a common buffer, more rigid SOPs were then used in the analyses (Figure 11.2). The flexible SOPs used in the initial stages of analysis are usually permitted as long as their details are well documented in the sample analysis records, are well controlled where possible (i.e., internal extraction positive controls, laboratory quality system in place, etc.), and in some cases, are approved by the submitting agency. It is important that documentation of laboratory actions, including technician's initials, date, and descriptions of deviations from SOPs be performed at the time they occur, or as soon as possible, rather than hours or days later.[4,5]

The development of test plans to accommodate requirements for rapid initiation of processing can be accomplished by using work sheets which are designed for rapid entry of pertinent information, such as the nature of the sample, quantity, biological threat agent(s) to be assayed, how the sample is to be extracted, etc. In many cases, portions of the test plan can be formulated before the samples arrive at the laboratory, as samples often arrive with advance notice, prior knowledge of their history, and an understanding of the testing required.

EFFECTS OF FIELD OPERATIONS ON LABORATORY TEST PLANS

The information obtained from the field, or from the agency delivering the sample, can have a direct bearing on decisions pertaining to the preliminary assessment, analytical test plan, and the type of laboratory capabilities required in the analyses (Figure 11.1). If a laboratory manager determines from the field information that the analyses requested are outside the laboratory's capability, he or she should immediately inform the submitting agency to consider alternatives. Otherwise, poorly performed analyses or limited capabilities of an analytical laboratory may be difficult to defend in court, thereby compromising the case and wasting valuable sample, time, and resources.

In cases where there is limited sample and limited analytical capability in the field, decisions must be made to determine the extent of testing to be performed on site before the sample is shipped to an analytical laboratory. In lieu of definitive assays at the site of an incident, there are proposals for first-responders to use simple assay kits to determine whether a powder or other material is a biological substance.[6–9] These proposed kits include tests for protein (found in most biological agents), pH (to determine whether the sample is in the biological range pH 6 to 8), and solubility (salts and sugars will produce a clear solution in aqueous fluids, whereas bacteria and spores will produce a turbid suspension). The downside of this approach is that valuable sample for subsequent laboratory analysis may be consumed and forensic evidence may be destroyed before the FBI and other investigational agencies have the opportunity to secure the situation. In other words, forensic testing may be hampered by contamination, alteration, or depletion of materials.[1,2]

EFFECTS OF FIELD CONDITIONS ON TEST PLANS

Developing test plans becomes more difficult when the threat agents being tested for are known to degrade or lose activity or viability. When such knowledge is unknown and the test plan is developed as per SOP, the analytical results may be either false negative or inconclusive. Unlike the dried *Bacillus anthracis* spores used in the threat letters in the fall of 2001 or ricin powder used in other incidents, many biological samples have a short half-life if not properly stabilized or handled in the laboratory. Hydrated materials, especially those containing an RNA virus or a protein toxin, may rapidly degrade, thereby reducing the analytical laboratory's ability to detect the agent. There have been reports of instances where samples tested positive with on-site assays in the field but were later found to be negative or below the detection level of the assay in the confirmatory laboratory. In such scenarios, questions arise; were the results from the field a false positive? Was there degradation of the sample over time while en route to the laboratory due to contaminating environmental organisms? Was there a problem with the confirmatory laboratory assay? Was the sample actually negative?

BIOLOGICAL THREAT CREDIBILITY, RISK ASSESSMENT, AND TEST PLAN PREPARATION

The credibility of an event has a direct bearing on the test plan and to what extent sample analysis is warranted (Figure 11.1). Field information from the submitting agency, contents of threat letters, sophistication of delivery devices, the nature of the threat material, clinical symptoms of exposed victims, communication with subject matter experts, and other intelligence all play key roles in risk assessment and in determining the credibility of the threat and the extent of testing to be outlined in a test plan. Certainly, a substance in an envelope containing a threat letter claiming to possess "anthrax" with the appearance of coffee grounds or tea leaves would not require the same extensive testing as that of a spray device attached to a reservoir tank. Risk assessment and determination of credibility also influence the management of potentially exposed victims, and may require the test plan to include preliminary analysis for specific agents to provide incident commanders and medical personnel with crucial information with which to make decisions.[6]

BIOLOGICAL AND PHYSICAL INTEGRITY OF A SAMPLE

Factors affecting biological integrity of agents/samples may be highly variable because of the diverse array of threat agent characteristics, environmental conditions, sample matrices, agent concentrations, adversary threat-agent preparation methods, and methods used in sample collection and transport. With respect to threat agent characteristics, agents range from labile protein toxins to stable mycotoxins, from labile RNA viruses to stable poxviruses, and from fastidious bacteria to very stable *B. anthracis* spores, all with varying degrees of purity. Less pure biological test samples may contain contaminating degradative enzymes, matrix components, overgrowing environmental organisms, etc., which may complicate isolation and identification. All of these variations make establishing a single sample-handling strategy to protect biological integrity very difficult. Problems with any one of these parameters can detract from the laboratory's ability to detect or identify the target threat agent while protecting forensic evidence. Attempts to stabilize one agent type against environmental factors and matrix effects may create adverse conditions for other agents. For example, freezing to stabilize a hydrated sample containing a protein toxin such as ricin against environmental proteases would in turn inactivate many viruses and bacteria. Gram-negative bacteria such as *Brucella* and *Francisella tularensis* are often stable for several days or weeks if dried in blood or proteinaceous materials, yet are unstable when dried in the absence of any protective matrices. Bacteria are subject also to their own autolysins, proteases, and nucleases and to the antimicrobial action of metabolites or compounds produced by environmental microflora. Such enzymatic or antimicrobial action could break down bacterial morphology, cause loss of viability, and alter antigenic properties, staining characteristics, and other traits required for detection or identification. In general, the primary enemies of sample integrity are time, adverse temperatures, contamination, and in certain cases moisture. The integrity of a sample with regard to degradation, viability, and its antigenic makeup is typically minimally affected during short time periods (minutes to a few hours); however, as this time lengthens, the sample may lose integrity, especially hydrated samples. In such cases, temperature becomes critical. Refrigeration or packing on wet ice slows the enzymatic degradation, but does not completely eliminate it; nor does it completely stop microbial growth of environmental organisms such as fungi, which are well known to grow on refrigerated foods. Therefore, hydrated materials typically require freezing if they must be refrigerated beyond 2–3 days. It is difficult to establish consistent policy for such decisions. A suggested approach for making decisions concerning storage of unknown biological materials is to treat the biothreat material as though it is powdered milk. When dry, powdered milk is stable and does not degrade when stored at room temperatures (RT) or lower. However when hydrated, milk spoils rapidly at RT, but lasts for several days refrigerated (depending on the level of initial contamination), but should then be frozen for long-term storage. If frozen, freeze-thaw cycles should be avoided and may require splitting samples or sample extracts into aliquots that are individually used for subsequent analyses. When the identity of the biothreat agent is known, one should use conditions appropriate for that agent or class of agents.

Stabilizing samples without altering the forensic evidence remains a critical consideration and requires research efforts in collaboration with law enforcement to determine how best to preserve a broad spectrum of biological agents while preserving forensic evidence. When one considers all the biological threat agents that are potentially available, it was fortuitous for forensic analyses that *B. anthracis* spores were used in the letters sent to the U.S. senators and the media in the fall of 2001. Spore powders are usually highly stable under desiccated storage conditions, and protecting the forensic evidence was much easier than if the material had been a more labile threat agent. In other words, there was minimal loss of viability (compared to what might have occurred with other infectious agents), samples were not degraded, and forensic evidence was not destroyed or its damage was minimal.

LEGAL CONCERNS FOR SAMPLE HANDLING AND DATA RECORDS

SAMPLE HANDLING

Matters of sample integrity are important forensic issues for both prosecution and defense lawyers. For prosecuting attorneys, the physical and biological integrity of samples to be analyzed become paramount not only for detecting and identifying biological threat agents, but also for analyzing forensic evidence. The defense lawyer's success in the courtroom may result from attempts to question the admissibility of analytical data based on poor sample handling. Questions may be posed regarding whether there was adequate separation of tasks in the analytical laboratory, which could have led to cross-contamination. These questions might include: Was there adequate security? Did unauthorized personnel have access to the samples? Was there adequate accountability (i.e., chain-of-custody, evidence storage, evidence in-processing)?[4,5,10] For example, defense attorneys may argue that samples brought into the laboratory may have become contaminated or that someone deliberately contaminated the sample in an attempt to incriminate their client. Therefore, it is important to determine those persons who need access to sample storage or work areas, and restrict access to just those individuals. Make a record of all individuals who have regular access, monitor their access, and have others from outside the laboratory sign a roster when it is necessary for them to enter the area. Access to samples must be restricted. There should be no exceptions.

With regard to tracking the location of the sample and access to it, it is imperative that chain-of-custody (COC) be established as a routine part of sample flow into the laboratory (Figure 11.1), preferably using approved COC forms provided by the submitting agency. However, in the absence of official standard forms during direct personal delivery of a sample to the laboratory, a simple printed statement of the transfer including a list of items, date, and time of transfer along with the printed names and signatures of both delivery and receiving personnel is recommended. The receiving laboratory personnel should only sign for what they can physically see during the transfer. Often, samples are delivered in biological safety packs by express delivery services such as Federal Express (FedEx®). In such cases, both the sender and the receiving laboratory should record bill of lading information and retain copies associated with the packaging. In those cases where a submitting agency directly delivers a sample to the laboratory, the identity of individual(s) making the delivery (photocopy of identification card) should be recorded and maintained with the records. Ideally, sample containers should have evidence tape applied to their lids and signed by an authorized representative of the submitting agency. A written record, and if possible a photograph, should document whether or not the evidence tape was intact upon delivery.[4] It is also recommended to have a Polaroid or digital camera available to photograph packages and contents. When using a digital camera, the image files should be in a format that cannot be tampered with.

It is recommended that a folder or envelope system be established into which all COC forms, shipping information (e.g., FedEx® bill of lading), communication logs, in-processing forms, etc. for an individual case are placed. Once the sample has been logged in, it is also recommended that a coded sample numbering system be used to give it unique, traceable, laboratory identification. This unique identification may be used to mask the original identification so that technicians cannot associate the sample with specific incidents that may be in the news or with samples that have a classified history. Any laboratory identification number given to an item of evidence should be cross-referenced to the original identifier provided by the submitting customer. If a laboratory information management system (LIMS) has been established, laboratory identification numbers or bar codes can be assigned at this time. It is also advisable to maintain an "Activities and Communications Log" as an aid in documenting occurrences, phone calls, and other communications. The advantage of the envelope or folder system is that if subpoenaed by the court or other legal entity, only the data for that case is released rather than bound notebooks with multiple cases being released. It is also preferred that the pages in the completed folder be numbered, and each page marked with the case number or laboratory reference number.

Another legal concern pertains to the potential or perceived potential for sample contamination. Great care must be used to protect the sample from other materials or threat agents that may be in the laboratory or that were associated with a previously analyzed sample. Observe the work area to determine whether it provides laboratory conditions that a court of law may question regarding whether a sample was cross-contaminated. Methods should be implemented as standard SOPs for demonstrating the lack of contamination from previous or adjacent activities. These methods may include culture of swabs from the work area, polymerase chain reaction (PCR) and analysis of liquids from air samplers or slit samplers within the work space or biological safety cabinets. It is preferable to protect a portion of the original sample from the testing area by submitting aliquots of a sample rather than introduce the entire sample into the testing area(s). By doing so, a second portion can be submitted either to the same testing laboratory or to another for confirmation if there are questions concerning whether the initial sample aliquot was possibly cross-contaminated. In those cases where it is difficult or impossible to divide a sample into portions for analysis, it may be necessary to extract the sample with an extraction buffer and to then supply each testing laboratory or area with a portion of the extract (Figures 11.1 and 11.2).

DATA HANDLING

Whether portions of the actual sample go into the testing areas or only portions of an extract, the transfer of these materials should be documented through an intra-laboratory COC system or LIMS. Such systems not only maintain COC, but also act as part of the tracking system for data retrieval from the testing areas and for documenting the final disposition of the sample material. Analytical data reported from the testing area should bear the sample identification code or LIMS designation and be stored in the sample's file. It is preferable for all data to be placed in the respective sample data folder. However, if there are original data remaining in the laboratory, there should be information provided in the sample folder or LIMS tracing the sample analysis to specific laboratory notebooks or computerized data files that are recorded as write-protected, read-only files. With any system, there will be errors in data and case records, and it is imperative that mistakes and corrections are clearly identifiable. Entry errors should not be erased or completely marked out, but rather the entry should be marked through with a single line, initialed, and dated. In that original data should be preserved, all notes should be saved. The data management system, whether as a hard copy or electronic, should be established or designed with the understanding that records may be subpoenaed years later. If the analytical laboratory chooses not to maintain the records for extended periods of time, then arrangements should be made with the respective law enforcement agencies to transfer the data and records to an approved location.

FINAL DISPOSITION

Also included in the records should be a record of the final disposition of the sample and materials associated with it. Often, laboratories receive a sample for analysis without clear guidance as to how the customer or submitting agency wants the sample disposed of after analysis. In such cases, the laboratory should request written instructions from the agency, detailing the final disposition of the sample. If the sample evidence is returned, COC must once again be documented (note: COC should have already been established) and recorded in the sample documents and the analytical laboratory's records. Note that sample return involves certain risks. In scenarios where an agency has requested an analytical laboratory test for specific agents and all the analyses are reported as negative or below the sensitivity of the analytical procedure, there are risks that other, undetermined threat agents may be present that were not tested for. Where possible, samples or items should be treated or sterilized by methods that are universal in their inactivation of biological compounds, e.g., 0.5% sodium hypochlorite bleach, dry heat (160°C for 2 h), or autoclaving. Where available, infectious agents may be inactivated with irradiation or vaporized hydrogen peroxide. Inactivation may damage forensic evidence, so the submitting agency should play an active role in the decision-making process as to how the sample will be treated. Regardless of how the material is treated, there should be documentation and certification of inactivation provided so that future laboratories or agencies involved will have assurance that their staff is not being exposed to an active biological threat. In some cases, the agency may want the sample or a portion of it sent to another analytical laboratory rather than to the submitting agency. Depending on the nature of the sample, special arrangements should be made to repack the sample for safe transport. In order to preserve all trace and latent evidence, the original packing materials may also be kept with the evidence (see the section below describing shipping regulations).

SAFETY ISSUES

The Centers for Disease Control Laboratory Response Network to Bioterrorism (LRN) recommends that samples to be analyzed for biological threat agents be processed within a certified Class II biological safety cabinet (BSC). Most diagnostic procedures deal with hydrated samples such as tissues and bodily fluids, and many biological samples from the environment are also hydrated, e.g., air collector fluids, water washings, etc. By the nature of such samples, procedures and protective equipment are usually adequate to handle most diagnostic samples safely within the aforementioned Class II BSC at biosafety level (BSL)-2 biocontainment levels.[11] However, certain forms of biological threat agents present unique problems with respect to their safe handling while protecting the forensic evidence. For example, powders can easily disperse in the air currents within a Class II biological safety cabinet, and in those cases where preserving forensic evidence is important, introducing liquid disinfectants to control contamination creates a different problem. If humidity alters the physical characteristics of the powder, future studies could be at risk. The difficulties in handling powders while protecting the safety of the technician, the properties of the sample, and forensic evidence require special conditions and may require higher levels of biocontainment (BSL-3 or BSL-4) than would otherwise be required for hydrated samples. This is especially true when dealing with certain nonclinical samples such as threat letters or with contaminated materials such as clothing, carpet, and items that do not fit easily and safely into a BSC.

If there is concern that using decontamination fluids will be detrimental to the forensic evidence, use only after the evidence is sealed. However, if it is impossible to prevent contact of the sample with disinfectants or to safely handle the sample, consider making special arrangements with the submitting agency to send the samples elsewhere. Where receipt of such samples is anticipated, establish protocols and obtain additional equipment (i.e., Class III BSC glove-boxes) to meet the requirements for protection of the sample and potential forensic evidence while safe-guarding laboratory personnel and property.

Other considerations put forth by the Centers for Disease Control and Prevention (CDC)[12] include: “Procedures requiring removal of items from a BSC, such as slides for microscopy, should follow published microbiological practices and precautions. When using a BSC, assure that the cabinet does not contain unnecessary items that will interfere with proper airflow and function. As for any procedure involving infectious materials, standard personal protective gear should be used, such as latex gloves and laboratory coats, or disposable over garments." Additional respiratory protection should also be considered with materials or analytical procedures determined to be potentially hazardous outside the BSC or when dealing with powders or items that do not fit well into a BSC. Once a biological agent has been identified, modifications in handling of samples can then be considered through a safety risk assessment of the set up.

To better address the safety issues, the following definitions of biosafety levels are provided[12] along with additional comments related to sample processing:

BIOSAFETY LEVEL 1 (BSL-1)

BSL-1 is suitable for work involving well-characterized agents not known to consistently cause disease in healthy adult humans, and of minimal potential hazard to laboratory personnel and the environment. The laboratory is not necessarily separated from the general traffic patterns in the building. Work is generally conducted on open bench tops using standard microbiological practices. Special containment equipment or facility design is neither required nor generally used. Laboratory personnel have specific training in the procedures conducted in the laboratory and are supervised by a scientist with general training in microbiology or a related science. Except for in-processing samples where samples are not exposed to the laboratory environment, samples may be received under this level of containment. Aside from the administrative aspects of in-processing sample packages, BSL-1 is not recommended for analysis of biological threat samples.

BIOSAFETY LEVEL 2 (BSL-2)

BSL-2 is similar to BSL-1 and is suitable for work involving agents of moderate potential hazard to personnel and the environment. It differs from BSL-1 in that laboratory personnel have specific training in handling etiologic agents and are directed by competent scientists, and access to the laboratory is limited when work is being conducted. All manipulations of biological threat materials are conducted in BSCs or other physical containment equipment.

Most hydrated diagnostic samples containing biological threat agents can be safely processed within a BSC at BSL-2.[11] However, great care must be used with finely divided, easily dispersed powders, and it is generally recommended that such samples be processed in a Class-III BSC (glove box) or be transferred to BSL-3 containment.

BIOSAFETY LEVEL 3 (BSL-3)

BSL-3 is preferred over BSL-2 for working with unknown biological threat samples in that many of these agents may cause serious or potentially lethal disease as a result of exposure by inhalation. Laboratory personnel have specific training in handling pathogenic and potentially lethal agents, and are supervised by competent scientists who are experienced in working with these agents. All procedures involving the manipulation of infectious materials are conducted within BSCs or other physical containment devices, or by personnel who wear appropriate personal-protective clothing and equipment. The laboratory should have special engineering and design features. Handling easily dispersed biological threat powders such as dried *B. anthracis* spores is especially hazardous. In addition to working within a Class II BSC, it is recommended that laboratory personnel wear double layers of protective gloves and disposable arm protectors to guard against contamination of laboratory apparel. It is also recommended that disposable protective gloves and arm protectors be either bleached or discarded before being removed from the BSC. If available, it is preferable to work within a Class III gloved BSC. Great care should be used to develop methods to safely remove materials from these cabinets while protecting the materials from moisture and decontaminating solutions, which may alter forensic evidence.

BIOSAFETY LEVEL 4 (BSL-4)

BSL-4 is required for work with dangerous and exotic agents that pose a high individual risk of aerosol-transmitted laboratory infections and life-threatening disease. It is rare that BSL-4 containment is required for processing most biological threat samples unless intelligence or the submitting agency can provide information that warrants such measures. Materials transferred into BSL-4 laboratories are difficult to remove and forensic analyses are more difficult to perform than at lower levels of biological containment.

SAMPLE SHIPPING REGULATIONS

Regulations regarding the transportation and transfer of biological agents are referenced in CDC websites and are summarized as follows. It is recommended that laboratories involved in shipment of biological threat agents or samples potentially containing such agents periodically check with CDC for regulatory changes. As throughout sample handling within the laboratory, final disposition of samples or their transfer to other locations requires continued COC documentation. If samples or evidence require inactivation through gamma-irradiation or other methods, a letter should accompany the shipment certifying that the material can be safely handled at the next location. If select agents are being sent, then compliance with governing regulations must be followed. The following describes some of these regulations.[12]

CDC/NIH REGULATIONS

Biological agents include infectious agents of humans, plants, and animals, as well as the toxins that may be produced by microbes and by genetic material potentially hazardous by itself or when introduced into a suitable vector. Etiologic agents and infectious substances are closely related terms that are found in the transfer and transportation regulations. Biological agents may exist as purified and concentrated cultures but may also be present in a variety of materials such as body fluids, tissues, soil samples, etc. Biological agents and the materials that are known or suspected to contain them are recognized by federal and state governments as hazardous materials, and their transportation and transfer are subject to regulatory control.

Transportation refers to the packaging and shipping of these materials by air, land, or sea, generally by a commercial conveyance. Transfer refers to the process of exchanging these materials between facilities.

TRANSPORTATION

Regulations on the transportation of biological agents ensure that the public and the workers in the transportation chain are protected from exposure to any agent that might be in the package. Protection is achieved through (a) the requirements for rigorous packaging that will withstand rough handling and contain all liquid material within the package without leakage to the outside, (b) appropriate labeling of the package with the biohazard symbol and other labels to alert the workers in the transportation chain to the hazardous contents of the package, (c) documentation of the hazardous contents of the package should such information be necessary in an emergency situation, and (d) training of workers in the transportation chain to familiarize them with the hazardous contents so as to be able to respond to emergency situations.

REGULATIONS

Public Health Service 42 CFR Part 72. Interstate Transportation of Etiologic Agents
This regulation is in revision to harmonize it with the other U.S. and international regulations. A copy of the current regulation may be obtained from the Internet at:

http://www.cdc.gov/od/ohs

Department of Transportation. 49 CFR Parts 171–178. Hazardous Materials Regulations

Applies to the shipment of both biological agents and diagnostic specimens. Information may be obtained from the Internet at:

http://www.dot.gov.rules.html

United States Postal Service. 39 CFR Part 111. Mailability of Etiologic Agents
Codified in the Domestic Mail Manual 124.38: Etiologic Agent Preparations. A copy of the Domestic Mail Manual may be obtained from the Government Printing Office by calling (202) 512-1800 or from the Internet at:

http://www.access.gpo.gov

Occupational Health and Safety Administration (OSHA). 29 CFR Part 1910.1030 Occupational Exposure to Bloodborne Pathogens. Provides minimal packaging and labeling requirements for transport of blood and body fluids within the laboratory and outside of it. Information may be obtained from your local OSHA office or from the Internet:

http://osha.gov

DANGEROUS GOODS REGULATIONS (DGR), INTERNATIONAL AIR TRANSPORT ASSOCIATION (IATA)

These regulations provide packaging and labeling requirements for infectious substances and materials, as well as clinical specimens that have a low probability of containing an infectious substance. These are the regulations followed by the airlines. These regulations are derived from the Committee of Experts on the Transport of Dangerous Goods, United Nations Secretariat, and the Technical Instructions for the Transport of Dangerous Goods by air which is provided by the International Civil Aviation Organization (ICAO). A copy of the DGR may be obtained by calling 1-800-716-6326 or through the Internet at:

http://www.iata.org or http://www.who.org

GENERAL PACKAGING REQUIREMENTS FOR TRANSPORT OF BIOLOGICAL AGENTS AND CLINICAL SPECIMENS

There are several commercial sources of packaging approved for transporting biological agents and clinical specimens. In general, a triple-layered system is required, including a primary receptacle, watertight secondary packaging, and durable outer packaging for a biological agent of human disease or materials that are known or suspected of containing them. This packaging requires the “Infectious Substance” label to be shown on the outside of the package and should be certified to meet rigorous performance tests as outlined in the Department of Transportation (DOT), the United States Postal System, the Public Health Service (PHS), and IATA regulations. Clinical specimens with a low probability of containing an infectious agent are also required to be “triple” packaged, but performance tests require only that the package shall not leak after a four-foot drop test. DOT, PHS, and IATA require a “diagnostic specimen” label on the outside of the package. DOT regulations require proper triple packaging if the sample travels over any public road or flies over United States territory, even if the sample is being delivered in person.

TRANSFER

Regulations on the transfer of biological agents are aimed at ensuring that the change in possession of biological materials is within the best interests of the public and the nation. These regulations require documentation of the personnel, facilities, and justification of need for the biological agent in the transfer process and subsequent approval of the transfer process by a federal authority. The following regulations fit in this category.

IMPORTATION OF ETIOLOGIC AGENTS OF HUMAN DISEASE

42 CFR Part 71 Foreign Quarantine. Section 71.54 Etiologic Agents, Hosts and Vectors This regulation requires an import permit from the CDC for import-ing etiologic agents of human disease and any materials, including live animals or insects, which may contain them. An application and information on importation permits may be obtained by calling 1-888-CDC-FAXX and entering document number 101000 or on the Internet at:

http://www.cdc.gov/od/ohs/biosfty/imprtper.html

IMPORTATION OF ETIOLOGIC AGENTS OF LIVESTOCK, POULTRY, AND OTHER ANIMAL DISEASES

9 CFR Parts 92, 94, 95, 96, 122 and 130 These regulations require an import permit from the U.S. Department of Agriculture (USDA), Animal and Plant Health Inspection Service (APHIS), Veterinary Services, to import or domestically transfer etiologic agents of livestock, poultry, other animals, and any materials that might contain these etiologic agents. Information may be obtained at (301) 734-3277, or from the Internet at:

http://aphisweb.aphis.usda.gov/ncie

IMPORTATION OF PLANT PESTS

7 CFR Part 330. Federal Plant Pest Regulations; General; Plant Pests; Soil; Stone and Quarry Products; Garbage This regulation requires a permit to import or domestically transfer a plant pest, plant biological agent, or any material that might contain them. Information can be obtained by calling (301) 734-3277 or through the Internet at:

http://www.aphis.usda.gov/ppq/permits/plantpest/index.html

TRANSFER OF SELECT BIOLOGICAL AGENTS OF HUMAN DISEASE

42 CFR Part 72.6 Additional Requirements for Facilities Transferring or Receiving Select Agents Facilities transferring or receiving select agents must be registered with the CDC, and each transfer of a select agent must be documented. Information may be obtained on the Internet at: http://www.cdc.gov/od/ohs/lrsat

EXPORT OF ETIOLOGIC AGENTS OF HUMANS, ANIMALS, PLANTS, AND RELATED MATERIALS

Department of Commerce (DOC). 15 CFR Parts 730 to 799 This regulation requires that exporters of a wide variety of etiologic agents of human, plant and animal diseases, including genetic material and products that might be used for culturing large amounts of agents, obtain an export license. Information may be obtained by calling the DOC Bureau of Export Administration at (202) 482-4811 or through the Internet at:
http://bxa.fedworld.gov or http://www.bxa.doc.gov

The CDC website provides illustrations for the packaging and labeling of infectious substances and clinical specimens in volumes of less than 50 milliliters in accordance with the provisions of subparagraph 72.3(a) of the regulation on Interstate Shipment of Etiologic Agents (42 CFR, Part 72). A revision is pending that may result in additional package labeling requirements, but this has not been issued in final form as of the publication of this fourth edition of BMBL.

For further information or to obtain updates on any provision of this regulation, contact:

Centers for Disease Control and Prevention
Attn: External Activities Program
Mail Stop F-05
1600 Clifton Road N.E.
Atlanta, GA 30333, USA
Telephone: (404) 639-4418
FAX: (404) 639-2294
Alternatively, go to the following website:
http://www.cdc.gov/od/ohs/biosfty/shipregs.htm

REFERENCES

1. Budowle B. (2003) Defining a new forensic discipline: Microbial forensics. *Laboratory Division of FBI*. Pub. No. 02–12.
2. Keim P. (2003) Microbial forensics: A scientific assessment. *Am Acad Microbiol*, Washington, DC. Available at http://www.asmusa.org/acasrc/Colloquia/ForensicsReport.pdf.
3. Budowle B, Schutzer SE, Einseln A, et al. Building microbial forensics as a response to bioterrorism. *Science* 2003;301(5641):1852–1853.
4. Standards Council of Canada (2001) Guidelines for the accreditation of forensic testing laboratories, Pub. No. CAN-P-1578, 270 Albert Street, Suite 200, Ottawa, Ontario, Canada.
5. ISO17025 2003. ISO-Standards International, Bottrells Lane, Chalfont St. Giles, Bucks, UK.
6. Franz DR. Suspicious white powder. Letter to the editor, *Homeland First Response* 2003;Vol. 1, Issue 3, pp. 9–10.

7. Hedgpeth D. Firms tap new market for homeland security: Companies working to develop products for U.S. government. *Washington Post*, May 8, 2003; p GZ25.
8. Brethauer T, Ezzell J, Franz D, Hagen M, Holt G, Hoehl R, Kerr L, Brett L, Mills B, Reid R, Sellito M, Schlegel R, Smith C. Suspicious white powder: Center for Domestic Preparedness develops inexpensive field test for ruling out anthrax spores. *Homeland First Response* 2003, July/August pp. 26–31.
9. Wade M, Campbell M, Rogers J, Coursey B, Niyogi K, Coughlin J, Emanuel P. Evaluation of an inexpensive field test for ruling out the presence of biological threat agents in suspicious powders. *Biosecurity Bioterrorism* 2005 (submitted).
10. Wheelis M. Investigating disease outbreaks under protocol to the biological and toxin weapons convention. *Emerg Infect Dis* 2000;6:595–996.
11. Henchal E, Teska J, Ludwig G, Shoemaker D, Ezzell J. Current laboratory methods for biological threat agent identification. *In* Marty A (ed). *Clinics in Laboratory Medicine*, Vol. 21, No. 3, Philadelphia: W.B. Saunders Company; 2001.
12. CDC.NIH, Biosafety in microbiological and biomedical laboratories. 4th edition. Richmond JY, McKinney RW (eds). Washington, DC: U.S. Government Printing Office; 1999.

CHAPTER 12

Forensic Genetic Analysis of Microorganisms: Overview of Some Important Technical Concepts and Selected Genetic Typing Methods

CHARLES L. COOKE JR.*
Formerly of Advanced Systems and Concepts Office, Defense Threat Reduction Agency, Fort Belvior, Virginia

INTRODUCTION

The scientific discipline of forensic microbiology is beginning to take shape. The rapid development of the combined disciplines of microbiology, molecular biology, forensics, and epidemiology is being driven by urgent national security and law enforcement needs, tragically underscored by the anthrax attacks perpetrated in the fall of 2001 and the dire threats currently posed by international terrorists and rogue regimes. Several forward thinkers in the federal law enforcement and national security communities had previously recognized the need to develop a robust capability in forensic microbiology given the perceived growing threat posed by bioterrorism and biological weapons in the hands of hostile nations. These individuals had been working to promote such a capability before the horrible events of September 11, 2001 and the subsequent mailborne anthrax attacks that were perpetrated against our country.[1,2] Prior to 9/11, the efforts to build a forensic capability for the analysis of microorganisms or their toxins were largely designed to foster interagency collaborations on research efforts and feasibility studies in order to

*This article solely represents the views of the author and does not necessarily reflect the views of the United States Government, the U.S. Department of Defense, or the Defense Threat Reduction Agency.

Microbial Forensics
Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

make maximal use of the available funding while advancing capabilities and the knowledge base. As mentioned, recent national security concerns have forced the issue and have lead to a much larger, better funded, and more intensely coordinated effort to develop a strong microbial forensic capability. Clearly the need for microbial forensic capabilities and the vigorous involvement and response by the Federal Government will foster the creation of a formidable new discipline in forensic science and microbiology. Such a strong forensic microbiology capability coupled with traditional "gumshoe" investigative techniques will greatly aid in attributing the source of a biological attack. Ultimately a reliable mechanism to attribute biological attacks to their source will aid law enforcement in prosecuting the criminals and provide the national security community with a valuable set of tools which will help shape the appropriate response should the perpetrator of a biological attack be an international terrorist network or rogue nation-state. Additionally, the mere knowledge that the United States has developed a highly dependable means to identify the perpetrator of a biological attack may help reduce the threat of such attacks in the future by groups or individuals for whom a desire for anonymity is paramount.

Technology of course will be a key component of forensic microbiology. Fortunately we are living during a time in which developments in molecular biology are occurring at a rapid pace, and many have referred to the latter half of the twentieth and beginning of the twenty-first centuries as being the "age of biology."[3,4] The rapidly expanding body of knowledge in the molecular biology of microbial pathogens can and will play an important role in the development of molecular tools for conducting investigations of crimes where forensic microbiology is required to fully exploit the evidence. Certainly the advances in molecular biology over the last 25 years or so have had a great impact on forensic analysis of human DNA samples.[5,6] Likewise the same will be true for forensic examination of microbial genetic material whether the pathogen is of a bacterial, viral, or fungal nature.

The purpose of this chapter is to provide an overview of some of the important technical concepts that have driven the development of genetic typing systems for microorganisms. Some selected typing methods will be discussed, along with some of the more obvious pros and cons of the approaches. Concluding this chapter is a discussion of the possible synergistic intersection between the tools of molecular epidemiology (and taxonomy) and the need to develop a robust set of microbial forensic genetic analysis techniques. The reader should also bear in mind that the author's intent is not to provide an exhaustive analysis of every typing system developed to date nor expound upon all the forensic issues that must be resolved surrounding such techniques. Rather the intent is to give those unfamiliar with these concepts and techniques a foundation upon which they can build as required.

USEFUL DEFINITIONS AND CONCEPTS

There are some terms and concepts that will make it easier to understand pathogen strain typing and the technologies that are being applied to this end. Characterization of microorganisms by strain typing represents an intersection between taxonomy and epidemiology, so it is useful to start the discussion from that point.

TAXONOMY AND EPIDEMIOLOGY

A system for the forensic analysis of microorganisms will incorporate concepts from the scientific disciplines of microbial taxonomy and modern infectious disease epidemiology. A system of microbial forensic analysis, however, will go well beyond the practices routinely applied in either of the aforementioned disciplines solely. Given the incorporation of concepts from microbial taxonomy and epidemiology, it is useful to briefly consider some points from each and how they are relevant to forensic investigation.

Microbial taxonomy strives to classify organisms into defined groups, develop appropriate nomenclature for those groups, and classify previously unknown organisms.[7] Forensic analyses are very similar in that forensic analyses involve comparing samples from known and questioned sources to establish the origin and identity of the sample with the highest degree of scientific certainty.[1] Interestingly, it is likely that as a system of microbial forensic analysis evolves, such a robust analytic regime may help to sort out one of the more controversial points in microbial taxonomy, the definition of a microbial species. As our knowledge of molecular biology and genomics of microorganisms increases, the definition of a microbial species is undergoing constant refinement and is a source of hot debate among microbiologists.[7–10] The controversy in the definition of microbial species arises from the discussion over which characteristics are the most evolutionarily relevant to base a classification upon.[7] This nomenclature problem becomes compounded when the word “strain” is added to the lexicon. Some isolates of similar bacteria, viruses, and fungi are often referred to as “strains.” For purposes of this chapter, consider the following definition as a frame of reference: a species consists of strains of common origin that are more similar to each other than they are to any other strain.[8] A strain therefore can be thought of as a subdivision of a species.[11] In a forensic investigation it may be necessary to go beyond the definition of strain to understand how strains may be related to one another.

Investigations in infectious disease epidemiology, which seek to relate a disease outbreak to a specific pathogen, have benefited greatly from the use of genetic fingerprinting approaches to characterize suspect pathogens isolated from patients.[12] As molecular biology has become increasingly more sophisticated, so too has our ability to fingerprint or "type" a strain of any particular microbial pathogen. This ability has improved the practice of investigations of disease outbreaks, because it allows the epidemiologist to distinguish the organism responsible for an outbreak from other similar strains that may also be present in the environment or microbiota but may not be responsible for the disease seen in the cases.[12]

GENETIC CONSIDERATIONS

Genetic typing systems are able to differentiate between similar microbial strains because of variations, referred to as polymorphisms, which are present in the genomes of the organisms being compared. However, microbial species can vary widely with respect to the level of polymorphisms that are found within a genome, and this can be a complicating factor in being able to distinguish between strains.

Eukaryotic organisms, which include all organisms except bacteria and viruses, possess large and complex genomes.[13] Eukaryotic genomes contain large quantities of repetitive, highly polymorphic DNA sequences.[13] This highly repetitive DNA is generally found in regions of the genome that separate genes rather than internal to genes. However, repetitive DNA also can be found in eukaryotic introns. Repetitive DNA consists of repeated units of nucleotides (A, G, C, and T) that can very in length and composition. The size of tandemly repeated sequence elements can vary from as few as one nucleotide to more than 80 nucleotides. Repeat elements can be homomeric (i.e. comprised of the same base) or heteromeric (comprised of different bases). These repeat elements are referred to as variable number tandem repeat (VNTR) elements. A subclass of VNTR elements, the short sequence repeat (SSR) element, is comprised of repeat elements that vary from two to seven bases in length.

The highly accurate forensic identification of individuals by DNA fingerprinting exploits the large amounts of polymorphic repetitive DNA in the human genome.[6] Unfortunately with the exception of perhaps fungal pathogens, bacterial and viral pathogens have much simpler genomes, which contain smaller amounts of repetitive DNA. However, most bacteria do contain SSR elements, and many viruses also contain repeat elements of one type or another.[14,15]

In order to develop robust identification regimes for organisms lacking a great deal of polymorphic repetitive sequences, it will likely be necessary to investigate the use of another type of polymorphic DNA marker, the single nucleotide polymorphism or SNP for short (pronounced "snip"). DNA sequences comprising a gene serve as a set of instructions that are ultimately translated into a protein, whereas other gene-associated DNA sequences are involved in regulatory functions. In the case of regulatory sequences and gene-coding sequences, the actual order of nucleotides (As, Gs, Cs, and Ts) that comprise the genetic code is important. Changing a nucleotide at a specific position in a gene can range from no effect, to enhancing a function or complete loss of function.[16] The nature of the change will determine how stable the SNP will become in the genome. For example a change in a coding region that produces a gene-product that provides the organism with some advantage will be more likely to be retained in the genome of the organism in successive generations. SNPs can also reside in noncoding regions. Comparative genomic sequencing to high confidence for multiple strains of a species will help identify which SNPs are best for use in typing. This is currently being explored for *Bacillus anthracis*, which is one of the more genetically monomorphic bacteria (i.e., gene sequences vary little between strains) and the causative agent of anthrax.[17]

RESTRICTION ENDONUCLEASES AND POLYMERASE CHAIN REACTION

Restriction endonucleases and polymerase chain reaction (PCR) are two crucial tools of molecular biology which are incorporated into a number of strain typing technologies. For the nonbiologists these terms can seem quite intimidating, but in reality the use of these tools is quite simple to understand.

Restriction endonucleases are simply enzymes (proteins) which cleave DNA at a set target sequence. Imagine a restriction enzyme simply as a pair of molecular scissors that will cut a piece of DNA at a specific place into fragments of varying sizes. The enzyme will cleave DNA at every point the specific target sequence occurs within the DNA. Restriction enzymes recognize target sequences that are 4, 6, or 8 bases in length, depending upon the enzyme used. Nearly all restriction enzymes recognize a unique sequence, although some recognize the same sequence but cleave the DNA in that sequence at a different position. For example the restriction enzyme BamH1 recognizes the sequence G*GATCC, with the asterisk indicating the position at which enzyme will cleave the DNA. These enzymes are only found in and isolated from bacteria.[18] Restriction enzymes likely arose as a defense mechanism against viruses that infect bacteria. The viral DNA entering the bacteria will be cut apart by the enzymes and rendered incapable of infecting the bacterium.[18] Bacterial DNA is protected from the action of its own enzymes.

Purified restriction enzymes are used in the laboratory to cut DNA into fragments the number and size of which depend on the position and distance separating each target sequence. Fragments of DNA can be separated in a porous sieving matrix by their size using an electric field to propel the DNA. DNA is negatively charged and will move towards the positive pole in an electric field. The process of separating DNA fragments in this manner is called electrophoresis.[19] When the current is turned off, the DNA fragments will remain separated in the porous matrix and can be visualized to reveal a pattern of bands. The resulting pattern will vary depending on the size and number of the fragments generated by the enzyme digestion. Multiple enzymes can be used in the same reaction digest to generate more complex patterns. In a sense, this process is an indirect measure of sequence variability. That is, "indirect" in the sense that only the number of recognition sequence found in the DNA is identified; restriction digestion does not provide any information about the intervening sequence. The ability to cut DNA at defined sequences is one of the critical discoveries that helped foster the development of recombinant DNA technology and the biotech revolution that followed.[19]

The discovery of the PCR was truly a milestone in molecular biology. Kary Mullis, the inventor of PCR, published his seminal paper describing the process of PCR in 1986.[20] Likewise the development of the PCR using a heat-stable enzyme had a profound impact upon molecular biology, ultimately allowing for the development of rapid automated analytical systems.[21] Simply, PCR is an *in vitro* molecular copying technique which can generate millions of copies of a specific target DNA sequence located anywhere in a genome. This process is conducted in the laboratory in a test tube; generally biochemical processes when carried out in a laboratory test tube are referred to as being done *in vitro*. As summarized in Figure 12.1, PCR has four critical requirements which are: (a) two synthetic oligonucleotide primers (each ~20 nucleotides) that are complimentary to regions on opposite DNA strands (DNA is a double-stranded molecule that can best be thought of as a ladder twisted to resemble a spiral staircase) which flank the target sequence. Oligonucleotide primers are short pieces of DNA which can be chemically synthesized to any desired sequence composition; (b) a target sequence of DNA, referred to as "a template" which is embedded in a larger stretch of DNA. Target sequences can be anywhere from 100 to 35,000 bases in length; (c) a thermostable DNA polymerase that can tolerate temperatures of 95°C or greater. Currently, the enzyme of choice is called Taq polymerase, named because it was isolated from the bacterium *Thermus aquaticus* which lives near undersea geothermal vents; and (d) the four building blocks of DNA, called deoxyribonucleotides, A, G, C, and T.[19] Before the steps of a PCR can be initiated, the reaction components consisting of a large excess of the primers, the template DNA, the *Taq* polymerase, the deoxyribonucleotides, and a special buffer solution are mixed together in a test tube which is then placed into a computer-

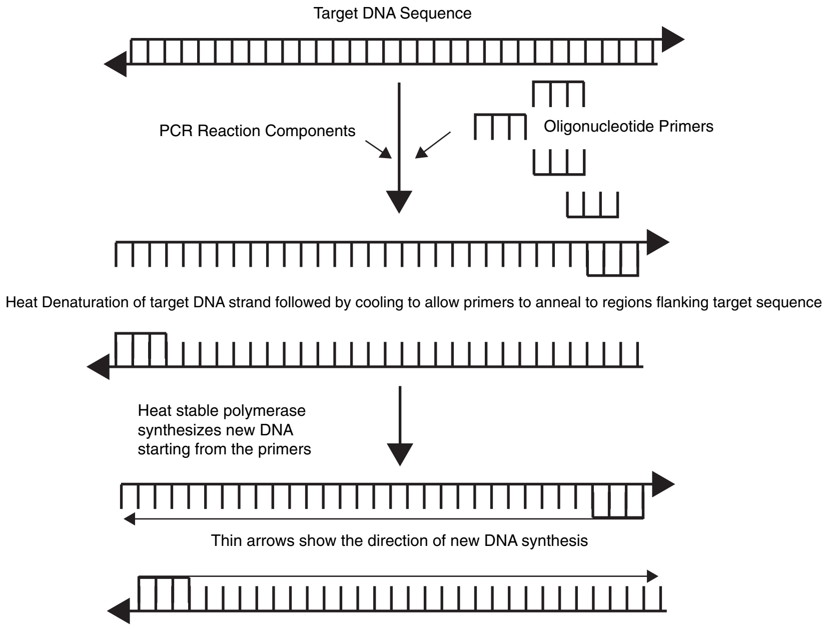

Target DNA Sequence
PCR Reaction Components
Oligonucleotide Primers
Heat Denaturation of target DNA strand followed by cooling to allow primers to anneal to regions flanking target sequence
Heat stable polymerase synthesizes new DNA starting from the primers
Thin arrows show the direction of new DNA synthesis

FIGURE 12.1 A schematic depicting the steps and components of a "single-plex" simple PCR.

controlled machine called a thermocycler. A thermocycler is essentially a programmable heating block into which the test tubes are placed. The machine will strictly control the temperature of the reaction inside the tubes for selectable amounts of time and through various cycles of heating and cooling. A PCR cycle consists of three steps. The steps in a PCR cycle are denaturation, renaturation, and synthesis. During the denaturation step, the temperature of the reaction is raised to 95°C for about 1 minute, which promotes the separation of the double-stranded template, or target, DNA into single strands. Next during the renaturation step, the reaction temperature is gradually cooled by the thermocycler to 55°C, during which time the primers (in vast excess relative to the target) will bind to their complimentary sequences in the template DNA. Finally, the temperature is raised to around 72°C, which is the optimal temperature for the *Taq* polymerase to function and begin synthesis of a new DNA strand using the target DNA as a template. A single PCR cycle lasts between 2–5 minutes, and the number of cycles in a PCR typically range from 25 to 40, depending on the assay. As the PCR reactions are subjected to more cycles, the amount of newly synthesized target DNA grows exponentially. It is in this manner that PCR is said to “amplify” the target sequence. Additionally reactions can be “multiplexed” by including additional pairs of primers for other target sequences such that multiple target sequences will be amplified simultaneously during each PCR cycle.[19]

OF DENDROGRAMS AND PHYLOGENETIC TREES

Although actual data from “strain typing” techniques will not be presented in this chapter, it is still useful to include a brief conceptual description of the format in which strain typing data are often presented, that is, the dendrogram or “tree diagram.” Dendrograms that show relationships among organisms based on evolutionarily relevant characteristics such as the variation in DNA sequence of a gene are referred to as phylogenetic trees. The organisms compared are placed on the tree diagram relationally to how divergent or similar the organisms are with regard to whatever characteristic(s) is being used to compare them. Some examples of compared characteristics include the number of nucleotide differences in a compared gene or DNA sequence, the presence or absence of bands generated during a restriction digest, or the presence or absence of amplified PCR products. The length of the lines on the tree and the distances between the branch points can be calculated by a variety of complicated mathematical algorithms. Textbooks have been written describing these algorithms, their application, and how to generate the correct type of tree diagram depending upon what type of characteristics are being compared.[22] However for those simply wishing to have a general understanding of what the data mean, the following is generally a good rule of thumb. The more branch points that separate two organisms on a tree diagram, the more different the two organisms are with regard to that characteristic(s). The converse is also true, that is, the fewer branch points that separate the organisms the higher the degree of similarity they have with regard to the characteristic(s) being compared. Organisms that are not separated by any branch points are thought to be indistinguishable with regard to the characteristic(s) being compared. As a caveat to this, the construction of these tree diagrams can vary for the same set of organisms depending upon which characteristic(s) is used for comparison and which algorithm is being used to calculate distance on the tree. This can lead to some controversy in estimated relationships if different characteristics vary wildly in evolutionary character. The careful observer should bear this in mind when “absolute” pronouncements are made about organism relatedness based on comparison of a single characteristic. Generally, the more characteristics that are compared which generate similar trees using different methods are likely better measures of relatedness between organisms.

MOLECULAR GENETIC TECHNIQUES FOR STRAIN TYPING

As mentioned in the introduction, this chapter is not intended to give a complete and exhaustive treatise on every type of technology used to genetically differentiate or "type" microorganisms. Rather, the intention is to give a broad survey of some of the more prevalent techniques present in the scientific literature, some representative examples of their use, and where possible discuss some of the pros and cons of employing a particular method. Whole-genome sequencing is a powerful technique, which will no doubt greatly aid forensic microbiology capabilities. Comparing the complete genomic sequence of two or more organisms is of course the most complete genetic-based comparison possible. Any and all genetic divergences and similarities between organisms can be readily discovered by this approach. However, the cost of employing this technology to the potentially large number of pathogen samples which could arise during the course of a forensic investigation renders the approach impractical for routine forensic analysis and similarly impractical for routine epidemiologic investigations. For example, the estimated cost to determine to high confidence the sequence of one complete genome consisting of ~5.2 million base pairs (less than 580 times smaller than the human genome) for the bacterium *Bacillus anthracis* was ~$140,000.[5] Sequencing 7–10 samples will cost well over $1,000,000. Whole-genome sequencing is a tool that currently is best used to identify highly variable regions of pathogen genomes that can be exploited for development of new genetic typing protocols and for forensic analysis.[5] Before whole-genome sequencing could be done as rapidly as it can be done today, and given the high cost, other less costly approaches were developed to genetically differentiate between related strains of microorganisms. These "strain-typing" techniques have served epidemiologists investigating the source of disease outbreaks and researchers studying evolutionary relationships between microorganisms. There is a considerable body of scientific literature describing these approaches and general acceptance of the utility of these techniques in the scientific community. Certainly this does not imply that there are never disagreements, but in general the techniques that will be discussed have found applications in both molecular epidemiology and research.

MULTILOCUS SEQUENCE TYPING

As mentioned, comparing the complete sequence of entire genomes will reveal any and all genetic differences between the organisms compared. Multilocus sequence typing (MLST) is a technique that can be used to compare microorganisms based on the DNA sequence of a set of genes rather than a single gene. This technique is usually used to compare microorganisms based on the sequences of conserved "housekeeping" genes (i.e., genes that are always expressed in the organism because they are essential for life functions).[23] This technique incorporates a PCR step to amplify the target regions to produce enough DNA for accurate sequencing.[23] Entire gene sequences can be examined, but generally only portions of the genes (some of which can be quite large) are sequenced and compared. MLST has been used in typing many bacterial pathogens of epidemiologic significance including *Neisseria meninigitidis*, *Listeria monocytogenes*, *Eschericia coli* O157:H7, and antibiotic-resistant strains of *Staphylococcus aureus*.[23–26]

Some advantages of MLST include a high degree of reproducibility between laboratories, the easy portability of sequence data between laboratories, and the ability to tailor the technique to examine nearly any set of genes for comparison. A major disadvantage of this technique becomes obvious when MLST is used to type genetically monomorphic microorganisms. For organisms such as *Bacillus anthracis* whose strains show very little genetic variability, it is very difficult to differentiate between isolates. In such strains, housekeeping genes and most other genes often show little or no variability, and, thus, other targets will need to be selected.[14,27] Another disadvantage to MLST arises from the possible introduction of bias when selecting genes to sequence, and even the region of a gene selected may not turn out to be the best choice, because the selected region may be less variable than other regions of the gene.

RESTRICTION FRAGMENT LENGTH POLYMORPHISM TYPING

The basic premise of this technology revolves around using restriction enzymes to digest a DNA sample into fragments and to separate the resulting fragments by size by electrophoresis to produce a characteristic pattern or "fingerprint." This pattern varies based on the number of times the enzyme recognition sequence occurs in the target DNA sequence and/or the number of intervening repetitive sequences contained within a generated fragment. Restriction fragment length polymorphism (RFLP) analyses can be used to examine restriction site polymorphisms in DNA targets of varied size. Indeed, RFLP typing was the first DNA fingerprinting method to be used to establish human identity.[6] For typing microorganisms, the differences in patterns between strains of the same organism are usually referred to as polymorphisms in this context.

RFLP analyses can be used to examine entire genomes. Following a genomic digest, a modified form of electrophoresis called pulsed-field gel electrophoresis (PFGE) separates the resultant fragments, some of which are very large in size. PFGE makes use of the fact that very large DNA fragments, which are usually not separated well using standard electrophoresis, can be resolved if the electric field is turned on and off at set intervals (pulsed).[13] Images of the electrophoresis patterns can be produced and compared for different strains and isolates of pathogens. The user community generally refers to this kind of typing simply as PFGE. The PFGE technique has been used successfully in molecular epidemiological investigations of foodborne and hospital-borne infection outbreaks for a wide number of pathogens. The relative utility of the approach has formed the basis of a sophisticated laboratory response network supported by the Centers for Disease Control and Prevention. This laboratory network has instituted standardized protocols, established computerized databases of PFGE patterns for outbreak strains, and has become well-integrated with the public health department labs in nearly every state.[28] It is truly an effort to be commended and one that is conceptually similar to the Combined DNA Index System (CODIS) database of human DNA fingerprints maintained by the FBI and U.S. Government.[6]

Ribotyping, another form of RFLP-based genetic typing, examines polymorphisms in restriction sites present in the ribosomal RNA gene. In eukaryotic organisms and bacteria, the ribosome is comprised of a protein-RNA complex crucial to the synthesis of proteins.[13] Ribosomal RNA (rRNA) genes are highly conserved but do show variability in some regions. Analyses of rRNA gene sequences have been used in evolutionary studies and in studies of microbial diversity. The Center for Microbial Ecology at Michigan State University maintains the “Ribosomal Database Project” with support from the National Science Foundation and the U.S. Department of Energy. The database holds over 64,600 bacterial ribosomal RNA gene sequences.[29] This enormous collection of sequences is an excellent source of information about bacterial diversity with respect to this gene, and helps to identify the most variable regions of the gene between and among species for use in developing typing protocols. As an example, ribotyping has been applied to investigate outbreaks of foodborne illness due to *Listeria moncytogenes*, *Chlostridium perfringens*, and *Salmonella enterica*.[30–32] Additionally, automated platforms have been developed to support this technique.[31] These automated systems help to reduce variability in the results and increase the speed and efficiency by which the analyses can be performed.[31]

While RFLP-based methods have been used with a good deal of success in epidemiological investigations and offer some advantages in the simplicity of the technique, these methods do suffer from some drawbacks. RFLP methods that do not incorporate a PCR step require larger quantitites of target DNA to employ, and restriction enzyme polymorphisms may have insufficient variation to distinguish between isolates of closely related strains. Ribotyping is not applicable to viral pathogens, because these parasites lack rRNA genes.

Other gene targets for RFLP analyses will need to be explored for viral pathogens. Some PFGE patterns from whole-genomic analyses can be very complex with many bands, which can make interpreting the patterns challenging.[11]

PCR-BASED GENETIC TYPING

There are a number of PCR-based typing systems which are in use, but for purposes of this discussion only two will be considered: amplified fragment length polymorphism (AFLP) and multilocus variable number tandem (MLVA) repeat analysis. AFLP was first described in 1995.[33] It is a powerful DNA fingerprinting technique that incorporates aspects of PCR and RFLP. In AFLP, restriction fragments from a total genomic digest (the enzymes employed are determined by the user) have short double-stranded oligonucleotides (adaptors) of defined sequence joined (ligated is the correct term) to both ends of the restriction fragment.[33] The adaptors can then serve as complimentary sequences for a defined set of PCR primers that are complimentary to the adaptor sequence and part of the restriction site sequence with some extension into the sequence flanking the restriction site. The fragments are then selectively amplified by PCR to produce a pattern of fragments of various sizes similar to the patterns generated by RFLP analysis. Incorporation of labeled primers which carry a fluorescent molecule (a “fluorescent tag”) has adapted the AFLP technique to run on automated DNA sequencing machines. Some examples of applications for the AFLP technique include typing geographically diverse collections of *Bacillus anthracis* isolates, studying the diversity of *Bacillus cereus* and *Bacillus thurengensis* isolates from Norwegian soil, and studying the genetic diversity of *Vibrio cholerae* (causative agent of cholera) isolates from the Chesapeake Bay.[34–38] AFLP is also being explored as a tool in bacterial taxonomy investigations, because of the power shown by AFLP to discriminate between highly related strains of the same species.[39] AFLP is advantageous because it can be used to type organisms for which very little or even no sequence data are available.

AFLP generates data that are, in a sense, binary, meaning the fragments are either present or they are not. Such “two-form” data in the lexicon of the geneticist is termed “bi-allelic” where alleles can be thought of as different forms of a genetic site or a gene. For pathogens that show very little genetic diversity, bi-allelic typing markers may be insufficient to discriminate between closely related isolates. More complex discriminatory markers with more than two forms may be required. VNTR elements are such markers, and they can have many different forms. The number of forms will vary according to the number of repeats that are found in the element. In VNTR analyses, sequences flanking the VNTR element are amplified using PCR primers, and the size of the fragments generated is then scored. Fragment sizes will vary depending upon how many repeats are present in the element.[40] Fluorescent-tagged PCR primers are used to facilitate automated analysis. By using different colored fluorescent tags and primers, the analyses can be multiplexed to score several VNTR elements at once, hence the name multilocus VNTR analysis (also known as MLVA).[40] This approach has been shown to be very powerful for discrimination between closely related isolates of organisms such as *Bacillus anthracis*, *Yersinia pestis*, and *Francisella tularensis*.[40–42] The power of the MLVA approach lies in the multi-allelic nature of the data produced compared to the bi-allelic nature of AFLP or RFLP data and the high mutation rate of VNTR elements. The drawback to the MLVA approach is that it cannot be used on an organism for which no sequence information is available. DNA sequencing (or hybridization assays) is required to identify possible VNTR locations (loci) in the genome of an organism.

VIRUSES AND FUNGI

Nearly all of the techniques discussed so far have been largely applied to the typing of bacterial pathogen strains. However, typing systems have been employed to examine viral pathogen strains and fungal pathogen strains. So it is worth discussing these pathogens as well. By and large the systems to genetically type viral pathogens and fungal pathogens are not as widespread in use as those implemented for bacterial species.

Viruses pose a unique set of problems for the development of typing regimes. However, these problems are not insurmountable. For example, many viruses use RNA instead of DNA as their genetic material, which renders direct DNA fingerprinting of viruses such as foot-and-mouth disease virus, influenza, or measles virus impossible. In order to employ RFLP-type methods on these viruses, the viral genetic material must first be converted to DNA and then amplified using a PCR step. This process, called reverse transcription (RT)-PCR, utilizes the enzyme reverse transcriptase to convert the viral RNA genome to DNA, which can then be amplified by PCR.[43] However, for viruses with smaller genomes the number of bands produced may be few. Viruses possessing small genomes (viral genomes can range in size from a few thousand bases to several hundred thousand) may best be typed by direct sequence analysis, both for cost and efficiency. Also, by using sequence analysis from many similar viral isolates, SNP loci can be identified for use as typing markers. In the near future as more and more viral isolates are sequenced, the development of rapid typing procedures should proceed at an accelerated pace.

Although fungi possess DNA as their genetic material, their genomes are more complex and larger than those of viruses or bacteria. With regard to typing pathogenic fungi, two techniques predominate: RFLP analysis, which has already been discussed, and a PCR-based technique called random amplified polymorphic DNA (RAPD).[44,45] RAPD has also been used with some success in typing bacterial strains.[1] RAPD uses a multitude of oligonucleotide primers, 10 nucleotides in length, consisting of random sequence composition to amplify fragments throughout the genome by PCR. Numerous primer sets must be tested to achieve the desired levels of variability in the datasets to produce meaningful discriminatory typing results.[44] The effectiveness of RAPD typing varies widely with primer sets and fungal species and is not optimal for all fungal species.[44] Finally, it can be expected that many of the identification assays used for eukaryotes, such as humans, are likely to be used on fungi.

CONCLUSIONS

While there are many genetic typing systems available to examine diversity among microorganisms and differentiate between closely related clinical isolates, having a typing system is not the same as having a robust, reliable forensic examination technique. Forensic discrimination and identification systems for microorganisms will require high confidence levels in the data such systems generate. Many of the typing systems discussed have variable applicability for different species within the same class of organism, with some being quite powerful for discrimination between closely related isolates and others being less informative. For any typing systems selected as possible forensic tools, numerous issues surrounding variability between species of microorganisms will need to be understood, error rates will need to be quantified, and any issues of reproducibility between laboratories must be resolved. Additionally, given the inherently high mutation rates of microorganisms, the effects of mutation and population biology on the typing data must be understood before high-confidence data can be generated. Validation of genetic typing techniques also needs to be carried out on panels of well-known and fully characterized samples before being applied to obtain the most information possible. It is also desirable to have more than one type of forensic genetic analysis to use on a sample, and it will require a great deal of systematic comparison studies to determine which techniques may be best for which organisms.

Some species of organisms may require sequencing of many representative genomes before typing systems can effectively exploit the variation. For some viral pathogens with small genomes, either full genomic sequencing or SNP analysis may become the typing system of choice.

Although not discussed herein, new technologies, such as microarrays, will certainly aid in the development of robust forensic genetic typing systems. Microarrays commonly called biochips consist of tens of thousands of oligonucleotides of known sequence arranged in a grid on a solid surface such as a glass chip.[46] Microarrays could be especially useful for interrogating samples for SNPs or VNTRs. These applications also will have to be evaluated before they can be fielded as a capability.

Overall current genetic typing systems for microorganism provide a good foundation that can be built upon to develop the highly reliable system of microbial forensic genetic analysis our nation requires. For some organisms, such as the highly mutation prone RNA viruses, the very nature of their biology may prevent achieving forensic confidence levels such as those achieved in human DNA fingerprinting. However with vigorous and well-funded research programs in microbial forensics, it is likely that our capabilities will become very powerful. Such potent microbial forensic capabilities will no doubt improve public health epidemiology, deter our enemies from using biological weapons, and ensure that those who remain undeterred are ultimately brought to justice.

REFERENCES

1. Lane SP, Beugelsdijk TJ, Patel CKN (eds). *Firepower in the Lab: Automation in the Fight Against Infectious Disease*. Washington, DC: Joseph Henry Press; 2001:203–213.
2. Falkenrath RA, Newman RD, Thayer BA (eds). *America's Achilles' Heel: Nuclear, Biological and Chemical Terrorism and Covert Attack*. Cambridge, Massachusetts: The MIT Press; 1999;261–336.
3. Brown D. "The Secret of Life: Untwisting Genetic Mysteries." *Washington Post*, September 25, 1993:B1.
4. Editorial. "Dealing with Dolly." *Washington Post*, May 22, 1997:A24.
5. Cummings CA, Relman DA. Microbial forensics-"cross examining pathogens." *Science* 2002;296:1976–1979.
6. Budowle B, Smith J, Moretti T, DiZinno J. *DNA Typing Protocols: Molecular Biology and Forensic Analysis*. Biotechniques Books, Eaton Publishing. 2000.
7. Van Belkum A, Struelens M, De Visser A, Verbrugh H, Tibayrenc M. Role of genomic typing in taxonomy, evolutionary genetics and microbial epidemiology. *Clin Microbiol Rev* 2001; 14:547–560.
8. Dijkshoorn L, Ursing B, Ursing J. Strain, clone and species: Comments on three basic concepts of bacteriology. *J Med Microbiol* 2000;49:397–401.
9. Lan R, Reeves P. When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends in Microbiol* 2001;9:419–424.
10. Fitch W, Brisse W, Stevens J, Tibayrenc M. Infectious diseases and the golden age of phylogenetics: An E-debate. *Infection, Genetics and Evolution* 2000;1:69–74.
11. Tenover P, Arbeit R, Goering R, Mickelsen P, Murray B, Persing D, Swaminathan B. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: Criteria for bacterial strain typing. *J Clin Microbiol* 1995;33:2233–2239.

12. Giesecke J. *Modern Infectious Disease Epidemiology*, 2nd ed. New York, NY: Arnold Publishing; 2002:160–175.
13. Lodish H, Berk A, Zipursky S, Matsudaira P, Baltimore D, Darnell J. *Mol Cell Biol*, 4th ed. New York, NY: W.H. Freeman; 2000.
14. Van Belkum A, Scherer S, Van Alphen L, Verbrugh H. Short-sequence DNA repeats in prokaryotic genomes. *Micro Mol Biol Rev* 1998;62:275–293.
15. Flint S, Enquist L, Krug R, Racaniello V, Skalka A. *Principles of Virology: Molecular Biology, Pathogenesis and Control*. Washington, DC: ASM Press; 2000.
16. Lewis R. SNPs as windows on evolution. *The Scientist* 2002;16:16–18.
17. Read T, Salzberg S, Pop M, Shumway M, Umayam L, Jiang L, Holtzapple E, Busch J, Smith K, Schupp J, Solomon D, Keim P, Fraser C. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* 2002;296:2028–2033.
18. Snyder L, Champness W. *Molecular Genetics of Bacteria*, 2nd ed. Washington, DC: ASM Press; 2003.
19. Glick B, Pasternak J. *Molecular Biotechnology: Principles and Applications of Recombinant DNA Technology*, 2nd ed. Washington, DC: ASM Press; 1998.
20. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. Cold Spring Harbor Symp. *Quant Bio* 1986;51:263–273.
21. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988;239:487–491.
22. Nei M, Kumar S. *Molecular Evolution and Phylogenetics*. New York, NY: Oxford University Press; 2000.
23. Maiden M, Bygraves J, Feil E, Morelli G, Russell J, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant D, Feavers I, Achtman M, Spratt B. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci* USA 1998;95:3140–3145.
24. Salcedo C, Arreaza L, Alcala B, De La Fuente L, Vazquez J. Development of a multilocus sequence typing method for analysis of *Listeria monocytogenes* clones. *J Clin Microbiol* 2003;41:757–762.
25. Noller A, McEllistrem M, Stine O, Morris J, Boxrud D, Dixon B, Harrison L. Multilocus sequence typing reveals lack of diversity among *Escherichia coli* O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. *J Clin Microbiol* 2003;41:675–679.
26. Grundmann H, Hori S, Enright M, Webster C, Tami A, Feil E, Pitt T. Determining the genetic structure of the natural population of *Staphlococcus aureus*: A comparison of multilocus sequence typing with pulsed-field gel electrophoresis, randomly amplified polymorphic DNA analysis and phage typing. *J Clin Microbiol* 2002;40:4544–4546.
27. Price L, Hugh-Jones M, Jackson P, Keim P. Genetic diversity in the protectice antigen gene of *Bacillus anthracis*. *J Bacteriol* 1999;181:2358–2362.
28. Swaminathan B, Barret T, Hunter S, Tauxe R, and the CDC PulseNet Task Force. PulseNet: The molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerg Infect Dis* 2001;7:382–389.
29. Cole J, Chai B, Marsh T, Farris R, Wang Q, Kulman S, Chandra S, McGarrell D, Schmidt T, Garrity G, Tiedje J. The Ribosomal Database Project (RDP-II): Previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res* 2003; 31:442–443.
30. Aarnisalo K, Autio T, Sjoberg A, Lunden J, Korkeala H, Suihko M. Typing of *Listeria monocytogenes* isolates originating from the food processing industry with automated ribotyping and pulsed-field gel electrophoresis. *J Food Prot* 2003;66:249–255.

31. Schalch B, Bader L, Schau H, Bergmann R, Rometsch A, Maydl G, Kessler S. Molecular typing of *Chlostridium perfringens* from a food-borne disease outbreak in a nursing home: Ribotyping versus pulsed-field gel electrophoresis. *J Clin Microbiol* 2003;41:892–895.
32. Clark C, Kruk T, Bryden L, Hirvi Y, Ahmed R, Rodgers F. Subtyping of Salmonella enterica serotype Enteritidis strains by manual and automated PstI-SphI ribotyping. *J Clin Microbiol* 2003;41:27–33.
33. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res* 1995;23: 4407–4414.
34. Keim P, Kalif A, Schupp J, Hill K, Travis S, Richmond D, Adair M, Hugh-Jones M, Kuske C, Jackson P. Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. *J Bacteriol* 1997;179:818–824.
35. Jackson P, Hill K, Laker M, Ticknor L, Keim P. Genetic comparison of *Bacillus anthracis* and its close relatives using amplified fragment length polymorphism and polymerase chain reaction analysis. *J Appl Microbiol* 1999;87:263–269.
36. Keim P, Price L, Klevytska A, Price L, Zinser G, Smith K, Hugh-Jones M, Okinaka R, Hill K, Jackson P. Molecular diversity in *Bacillus anthracis*. *J Appl Microbiol* 1999;87:215–217.
37. Ticknor L, Koltso A, Hill K, Keim P, Laker M, Tonks M, Jackson P. Fluorescent amplified fragment length polymorphism analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* soil isolates. *Appl Environ Microbiol* 2001;67:4863–4873.
38. Jiang S, Louis V, Choopun N, Sharma A, Huq A, Colwell R. Genetic diversity of *Vibrio cholerae* in Chesapeake Bay determined by amplified fragment length polymorphism fingerprinting. *Appl Environ Microbiol* 2000;66:140–147.
39. Coopman J, Huys G, Swings J, Bleeker M, Vos P, Zabeau M, Kersters K. Evaluation of the DNA fingerprinting method AFLP as an new tool in bacterial taxonomy. *Microbiol* 1996; 142:1881–1893.
40. Keim P, Price L, Klevytska A, Smith K, Schupp J, Okinaka R, Jackson P, Hugh-Jones M. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol* 2000;182:2928–2936.
41. Klevytska A, Price L, Schupp J, Worsham P, Wong J, Keim P. Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. *J Clin Microbiol* 2001;39:3179–3185.
42. Farlow J, Smith K, Wong J, Abrams M, Lytle M, Keim P. Francisella tularensis strain typing using multiple-locus variable-number tandem repeat analysis. *J Clin Microbiol* 2001;39: 3186–3192.
43. Arens M. Methods for subtyping and molecular comparison of human viral genomes. *Clin Microbiol Rev* 1999;12:612–626.
44. Sol D. The ins and outs of DNA fingerprinting the infectious fungi. *Clin Microbiol Rev* 2000; 13:332–370.
45. Taylor J, Geiser D, Burt A, Koufopanou V. The evolutionary biology and population genetics underlying fungal strain typing. *Clin Microbiol Rev* 1999;12:126–146.
46. Constans A. The state of the microarray. *The Scientist* 2003;17:34–39.

CHAPTER 13

Non-DNA Methods for Biological Signatures

Charlene M. Schaldach,[1]* Graham Bench,[2] James J. DeYoreo,[1]
Tony Esposito,[1] David P. Fergenson,[1] James Ferreira,[3]
Eric Gard,[1] Patrick Grant,[2] Christopher Hollars,[1]
Joanne Horn,[4] Thomas Huser,[1] Michaele Kashgarian,[2]
John Knezovich,[2] Stephen M. Lane,[1] Alexander J. Malkin,[1]
Maurice Pitesky,[1] Chad Talley,[1] Herb J. Tobias,[1] Bruce Woods,[1]
Kuang-Jen Wu,[3] and Stephan P. Velsko[1,5]

Lawrence Livermore National Laboratory:
[1]*BioSecurity and Nanosciences Laboratory, Chemistry and Materials Science Directorate*
[2]*Center for Accelerator Mass Spectrometry, Energy and Environment Directorate*
[3]*Materials Science and Technology Division, Chemistry and Materials Science Directorate*
[4]*Energy and Environment Directorate*
[5]*Nonproliferation, Arms Control, and International Security Directorate*
**Corresponding author.*

The study of bioweapon manufacturing signatures is a multidisciplinary effort to elucidate the molecular and morphological signatures or "fingerprints" of microbial entities such as bacteria, viruses, and spores. Our goal is to determine how, when, and where these agents were produced, addressing key issues such as the sophistication and knowledge level of the producer, the geographical origin of the component materials used to produce the agent, and the time of manufacture relative to release or discovery. Of great importance are techniques that can analyze individual agent particles and thus are applicable to trace samples that may contain only a few agent particles, or samples that are contaminated by debris unrelated to the agent or its production. Observations are made at the single-organism level by employing high-resolution, high-sensitivity techniques such as atomic force microscopy, Raman spectroscopy, time-of-flight mass spectrometry, and accelerator mass spectrometry, which directly measure microbial intrinsic physical properties such as shape, size, mass, molecular vibrations, fluorescence, and isotope composition. Results from these analyses can be correlated with known growth

Microbial Forensics
Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

and processing steps. Ultimately, these disparate data types (e.g., images, spectra, and text) will be combined, producing a signature library that correlates morphological and compositional features with agent manufacturing methods to determine an organism's unique biological signature.

This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory, under Contract No. W-7405-Eng-48.

INTRODUCTION

According to Webster's Dictionary, a signature is "a distinguishing or identifying mark." Signatures are commonly used in the process of "attribution," defined by Webster's as "the process of ascribing (an event) to someone or something." A common signature is our name signed at the bottom of a letter; retinal scans and palm prints are becoming more common. This chapter is concerned with methods that can determine chemical or structural features of biological agent particles that are signatures of particular methods of growth and post-growth processing (often referred to as "weaponization"). The detection of these signatures in a sample of a bioweapon (BW) agent can aid the attribution effort by indicating the level of sophistication of the producer, his access to particular types of agent weaponization information, and the likelihood that he could produce—or has produced—the material at a significant scale, and by providing essential sample matching data for ascertaining a putative relationship with other samples obtained in other venues. In addition, it may be possible to identify at least some of the materials used in the manufacturing process, correlate them with reference materials that have known geographic or temporal provenance, and determine the date of manufacture of the agent to within months or weeks.

"Signature" differs from "forensic." Webster's defines "forensic" as "belonging to courts of judicature" and "used in legal proceedings." The ultimate goal of forensic sciences is to find evidence or information that will either incriminate or exclude a subject, enabling arguments in court that can prove to a reasonable degree of certainty that the crime was or was not committed by the accused. As such, many signatures comprise a forensic study. One signature alone is most often insufficient to convincingly prove guilt or innocence. It is the great preponderance of evidence, rather than the "smoking gun," that most commonly convicts a criminal. While a salient goal of the attribution effort is to find evidence that will enable courtroom arguments to be made, it should be noted that this is not the only utility of manufacturing signatures. Nonforensic utility includes delimiting the scope of an investigation, providing information for threat assessment (e.g., in the case of an extortion attempt where a BW agent is sent to law enforcement officials to demonstrate the credibility of the threat), and in informing decisions for military counterproliferation actions.

An example of the use of biologicals in forensic science is DNA, amplified by the polymerase chain reaction (PCR) technique, legally admissible in court as evidence. DNA evidence has been used successfully in court to convict or clear people of crimes because each person's DNA is unique. That same characteristic uniqueness of DNA has not been established for bioweapon agents. A BW agent may be obtained from an environmental source deliberately chosen to mislead investigators or, more simply, many people and laboratories may possess virtually genetically identical biological agents, thus reducing the uniqueness factor of DNA and establishing a requirement for other corroborating signatures to link the bioagent to the perpetrator of the biocrime. As we shall see, there are other powerful techniques which can be employed to uncover these signatures.

One of the basics of criminalistics is Locard's Exchange Principle:[1] When any two objects come into contact with each other, there is always a transfer from one object to the other. When you walk into a room, you "change" the room, and the room may change you. Enter, turn around, walk out without touching anything; you will still have left something behind: a grain of sand from your shoe, a fiber from your car, a hair belonging to you or someone you touched.

Manufacturing signatures can be thought of as a manifestation of that basic tenet of forensic science. While the person producing the agent does not necessarily leave hair behind, he or she may leave other clues: starting material, processing or growth media components. Just like a coroner at an autopsy, many things may provide a clue—nucleic acids, proteins, lipids, metabolites, crystals, surface charge, trace elements, elemental compositions, isotope ratios, and morphology. A simple example is the isotopic composition of microbes, which approximates that of the culture medium, and can be related to the isotopic composition of the water and nutrients used to grow them. In addition, culture conditions will always leave some mark on the structure and composition of an organism through the metabolic influence of nutrient composition, temperature, and aeration conditions. Metabolites resident in the culture medium are very likely to accompany the organisms when they are harvested and concentrated, and can specifically adhere to the outer surfaces of cells, spores, or virions.

In our attempts to identify these clues, we focus on those signatures that are *unalterable regardless of attempts to wash or prepare* the bioweapon to remove all traces of growth or processing methods. Therefore, in addition to looking at the bulk sample, we are investigating the *surface* as well as the overall *composition* of *individual* biological agents (e.g., a single bacterial

TABLE 13.1 Advantages and disadvantages of detection methodologies

Method	Advantages	Disadvantages
DNA	Unique identification of species and strains of organisms Admissible in court	Must have target sequences beforehand Requires amplification (PCR)[a] Provides no information on growth or processing methods
Non-DNA: indirect detection	Wide-spread use Large number of antibodies available Standard detection methods available Can be relatively inexpensive and portable	Must have antibodies beforehand No information if antibody does not bind to target False positive if antibody binds to the wrong target
Non-DNA: direct detection[b]	Always provides information: does not depend on prior knowledge of target No antibody required No target amplification (copying) is required Measure intrinsic properties of target (shape, elemental composition, vibrations)	Requires high-sensitivity, high-resolution techniques Signatures libraries used for identification Instruments can be expensive and relatively nonportable

[a] Polymerase chain reaction (PCR) is used to amplify (make thousands of copies of) a piece of DNA.

[b] See Table 13.2 for methods.

spore), searching for chemicals or signals which can uniquely identify **how, when, and where** the sample was made. We are correlating these signatures with known growth and processing steps. It should be recognized that DNA analysis provides no information regarding these issues. Many clues exist—we need to apply multiple high-resolution techniques to find them.

Before discussing specific non-DNA analyses, a brief explanation as to how our approach fits into the overall microbial forensics picture is in order (see Table 13.1). DNA is receiving a lot of attention as a signature of an organism, as described amply elsewhere in this book. However, in order to use DNA, PCR must be used to greatly increase its quantity by making copies (called "amplification") so that the DNA can be detected by today's techniques. Therefore, primers must be determined beforehand, that is, *a priori*. These primers may be specific for an organism, with the risk that the chosen primers will not be unique for the given organism, or the primers may be more generic, as in the case of amplified fragment length polymorphism (AFLP).[2] For example, if the primers are chosen to be unique to a given bacterial strain, they will only find that specific organism; other primers must be chosen to find other specific organisms. If the primer does not find the agent (i.e., amplify the DNA for detection), no information is learned beyond the fact that that specific organism is not there. If a more generic approach to generating DNA primers is taken, the specificity or certainty of identification of the organism is diminished.[3,4] Furthermore, a commonly used region for primer development (e.g., 16S rRNA genes) could be easily mutated such that the agent would escape detection. With either approach to DNA primers, no information addressing the method of preparation is obtained.

For *non-DNA* constituents, which themselves are not amenable to amplification, there exist *indirect* and *direct methods* for detection and identification (see Table 13.1). A major group of indirect detection methods employed in the present day is the immunoassay. These assays use an antibody to bind the target (e.g., toxin or protein), and the antibody-target complex is then detected by any of several means such as Western blot, immunohistochemistry, and flow cytometry. Similar to PCR primers, the antibodies must be generated *a priori* and, in addition, the antibody must bind to the target molecule in the particular environment that is required (i.e., tissue, gel, solution). If the antibody does not bind for any reason (e.g., the target is not there or the antibody could not find the target), no information is derived from the test, as was the case for DNA (PCR).

The antibodies which bind to a target molecule are themselves "tagged" with a fluorescent molecule so that when the antibodies attach to the target, that complex (antibody + target) can be detected. What is really being detected or measured is the fluorescent antibody—we assume it bound to the correct target. But how do we know that what really bound to the antibody is what we wanted? This can be more challenging than it may seem. For example, an antibody was believed to bind an organism, only to find that the antibody's actual target (epitope) was a media component stuck on the surface of the organism.[5] Furthermore, the fluorescent tags attached to antibodies are known to bleach after only a short time, no longer emitting any detection signal.

What is needed are *direct methods* which can fingerprint the non-DNA molecule itself by measuring the molecule's intrinsic properties (Table 13.2). These methods would be able to detect both known and unknown molecules, a major advantage in our changing world where genetic engineering is commonplace. The techniques must have the sensitivity and resolution to detect and identify single molecules or spores. PCR amplification and immunoassays have the disadvantage of requiring prior knowledge of the target (i.e., one only finds what one seeks); information about unknown molecules is lost. With direct methods, a known target can be identified through use of signature libraries, and an unknown entity detected and stored for further analysis. *In this way, all detected compounds provide information whether they are identified or not. No information is lost.*

TABLE 13.2 Summary of analytical techniques applied to the analysis of bacterial spores. Indications of which intrinsic properties are measured are included. Bulk refers to analysis from an entire spore, surface refers to measurements from only the surface of a single spore

		Images		Elements		Vibrations/ rotations		Masses	
Abbreviation	Name of technique	Bulk	Surface	Bulk	Surface	Bulk	Surface	Bulk	Surface
SEM/EDX	Scanning electron microscopy with energy dispersive X-ray microanalysis	x	x	x					
AFM	Atomic force microscopy[a]		x						
	Raman spectroscopy					x			
SERS	Surface-enhanced Raman spectroscopy[b]						x		
BAMS	Bioaerosol time-of-flight mass spectrometry[c]							x	
ToF-SIMS	Time-of-flight secondary ion mass spectrometry[d]				x				x
AMS	Accelerator mass spectrometry[e]							x	
PIXE	Particle-induced X-ray emission[f]			x					
STIM	Scanning transmission ion microscopy[g]							x	

[a] AFM with carbon nanotube tips can generate images with ~20 Å resolution.

[b] SERS utilizes nonfunctionalized nanoparticles; no need for antibodies and nanoparticles do not photobleach.

[c] BAMS is real-time, reagentless, and uses only electricity.

[d] ToF-SIMS analyzes 1 nm layer with 150 nm spot resolution and ppm sensitivity.

[e] AMS can quantify attomoles (i.e., 10–18 moles) of ^{14}C in samples of size <1 mg at a precision of better than 1%.

[f] PIXE provides quantitative analyses of element distributions with elemental sensitivities that can approach 0.1 mg/kg.

[g] STIM accurately quantifies sample mass without specific standards.

INTRINSIC MOLECULAR PROPERTIES: FINGERPRINTS

Just as people possess unique biometrics such as retinal scans and fingerprints, biological entities also have "fingerprints." Distinct *shapes and sizes* of molecules and other targets (e.g., spores) are the most obvious; both the overall bulk shape as well as detailed surface features are characteristic fingerprints amenable to high-resolution imaging. *Molecular vibrations and rotations*, created by forces unique to interactions between atoms comprising the molecules, are also specific fingerprints and can be directly measured by several techniques. Electrons in molecules also provide signatures: Molecules *absorb* UV light (electrons excited to higher energy states) at specific wavelengths characteristic of their molecular composition. They also *emit* or give off absorbed light (fluoresce) at characteristic frequencies (electrons return to a lower energy state and emit photons), providing another fingerprint of the molecule. When excited at high energies, electrons from the inner shells of atoms (closest to the nucleus of an atom) can be ejected. When other electrons fall back into the vacant site thus created, they do so simultaneous with emitting a characteristic *X-ray* photon. Individual atoms, molecule fragments, and entire molecules have *masses* which are characteristic fingerprints; importantly, *isotopic ratios* (same element, different mass) may indicate where the growth ingredients are from geographically.

The techniques to analyze these intrinsic molecular properties are well established. It is the application of the technologies to the problem of single-cell microbial forensics that is in its infancy and rapidly evolving. The intent of this chapter is to illustrate the potential of employing these technologies to examine individual bacterial spores, describing what may possibly become routine biological signature analysis in the future.

A MULTIDISCIPLINARY APPROACH TO BIOLOGICAL SIGNATURES

Achieving a sound scientific basis for correlating manufacturing processes to measurable intrinsic signatures will require a unique, multidisciplinary effort, ranging from information collection and analysis to experimental microbiology, analytical chemistry, physics, and state-of-the-art instrumental materials characterization. The Intelligence Community is a source of information about BW manufacturing processes associated with foreign programs. In addition, much of the information that is available to terrorists comes from a variety of unclassified sources such as underground cookbooks, Internet recipes, open scientific literature, and occasional releases of sensitive information in news-media reports. Manufacturing methods can be gleaned from these sources and used in simulation experiments, while remaining cognizant of relevant differences between the biological simulant and the actual bioagent.

We are employing a variety of techniques in the analysis of bioagents themselves, specifically bacterial spores. A summary of the analytical techniques to be discussed in this chapter is given in Table 13.2. The list is not meant to be exhaustive, but represents a sampling of the forward-looking capabilities we have in a single location to focus on this challenging problem of biological signatures; experimental data are shown to illustrate the application to the analysis of bacterial spores. All of these techniques, with the exception of accelerator mass spectrometry, can analyze a single biological entity such as a *single bacterial spore*, thereby obviating the need for large quantities of sample material and ensuring that our measurements are from the particle itself and not from contaminating matrices. We are working toward a common sample analysis platform whereby future analyses will be performed as much as possible on the *same exact spore*. Comparison of measurements across single particles will be important for establishing the precision and accuracy of inferences that can be made from such data. A brief description of each method (see Table 13.2) follows.

OVERVIEW OF CHARACTERIZATION TECHNIQUES

Scanning electron microscopy (SEM) provides excellent **images** of the spore surface. Low beam energies (approximately 1.2 keV) give images of the outer surface detail, while higher beam energies (approximately 15 keV) penetrate the outer exosporium and image the inner coats of the spore. When combined with **energy dispersive X-ray microanalysis** (EDX), the elemental composition can be determined on the exact same imaged sample, providing a direct connection between the image and its elemental composition. **Atomic force microscopy** (AFM) allows very-high-resolution imaging of the sample surface, scanning a wide range of fields from 20 nm (0.02 μm) to 150 μm. Resolutions on the order of tens of nanometers, with height resolution of 0.1 nm, are routine; by utilizing carbon nanotubes as imaging tips, image resolutions of less than 1 nm can be achieved. The major advantage of AFM is its ability to image directly in a fluid.

Raman and **surface-enhanced Raman spectroscopy** (SERS) probe molecular *bond vibrations and rotations* to produce characteristic spectra. Here, electrons in the molecules are excited from a vibrational level to a "virtual level" lying above the ground state vibrational levels but below the excited electron states and return to a different vibrational level, emitting a characteristic photon indicative of the "Raman shift" as they do so. Raman effects have a low "cross-section" or probability of occurring; however, by placing the molecule near a gold or silver surface, the Raman signal is enhanced by factors as large as 10^{15}! This effect is highly local (as is the electric field) and hence can be taken advantage of to probe surfaces of biological agents such as spores. Protein secondary structure can be determined by its characteristic Raman peaks; by adsorbing gold or silver nanoparticles (~80 nm in diameter) to the surface of a biological agent, only the proteins or other molecules in the local vicinity of the nanoparticle (of order 10 nm) are analyzed.

In general, mass spectrometry measures the *masses* (i.e., mass-to-charge ratios) of ions and molecules. There are several types of mass spectrometers and a wide variety of sample introduction and ionization methods. **Bio-Aerosol Mass Spectrometry (BAMS)** is a novel real-time technique for the rapid identification of individual bio-aerosol particles using mass spectrometry. It is the hybrid of a relatively new method of mass spectrometric sample introduction and a novel method of analyzing the resulting data in real-time. BAMS can operate autonomously, consumes only electricity, and is unique in that it allows for the reagentless analysis of a complex environmental sample without any prior sample preparation or discrimination. It is currently capable of detecting and identifying individual cells sampled directly from the air within an aerosol of many background materials in real-time with no reagents. *Bacillus* spores have been successfully characterized and can be efficiently distinguished from fungal spores, vegetative bacteria, and many other biological and nonbiological background materials in real-time.

Time-of-flight secondary ion mass spectrometry (ToF-SIMS) also separates molecular fragments by their size (and, hence, their "time of flight"), but this instrument utilizes an ion beam rather than a laser to generate those fragments. Because it uses an ion beam, the energy can be lowered so as to analyze approximately one atomic layer at a time. The ToF-SIMS captures elemental data (to generate chemical maps) as well as molecular fragments (to generate mass spectra), and does so on a depth-dependent basis (to generate a depth profile).

A third type of mass spectrometry, **accelerator mass spectrometry (AMS)**, is the most sensitive of all techniques included here. AMS combines mass spectrometry and nuclear detection to measure the concentration of an isotope in a sample. The accelerator is required because the molecules to be broken are only broken at million-electron Volt (MeV) energies, in contrast to the kilo-electron Volt (keV) energies normally used in mass spectrometry. For biological samples, ^{14}C measurements are made by first converting the sample to graphite, followed by ionization and subsequent acceleration to finally become a 35 MeV $^{14}C^{4+}$ ion. These ions are resolved by high-energy mass spectrometry and, ultimately, can be quantified in the range of attomoles (i.e., 10^{-18} moles) of ^{14}C in samples of size less than 1 mg at a precision of better than 1%.

Nuclear microscopy is a combination of **particle-induced X-ray emission (PIXE)** and **scanning transmission ion microscopy (STIM)**. PIXE uses characteristic X-rays to quantitatively map simultaneous element distributions (for elements greater than sodium) within microscopic regions of a sample, and STIM detects the energy loss of accelerated ions which pass through the sample in the exact same sample region. This loss of energy enables the mass of the sample to be accurately quantified. This technique is relatively nondestructive to the sample, allowing further analysis on the same exact components.

Having described these techniques by the molecular intrinsic properties they measure, we next look at the granularity or spatial resolution each technique can achieve on a sample. Accelerator mass spectrometry reduces the *entire sample* (less than 1 mg) to carbon before performing the analysis and provides ^{14}C measurements on the bulk sample. Similarly, scanning electron microscopy is useful for characterizing the overall sample through image as well as elemental data. Several of these techniques, however, can analyze a single biological entity such as *a (whole) spore*. Raman spectroscopy, bioaerosol time-of-flight mass spectrometry, nuclear microscopy, and SEM with energy-dispersive X-ray analysis, while examining just a single spore, derive their data from the signal generated by the whole or entire spore. By comparison, atomic force microscopy, surface-enhanced Raman spectroscopy, and time-of-flight secondary ion mass spectrometry are able to analyze *only the surface* of a single spore.

By using a combination of these complementary techniques, *we can investigate signatures of sample growth, processing, geolocation, and chronometry*. The biological signatures will be in the form of images, Raman spectra indicative of molecular composition, and elemental composition due to protein fragment and elemental masses. In addition, this information can be determined on the basis of its spatial distribution in the sample (spore).

In the following paragraphs, we describe each of these analytical techniques in more detail. First, high-resolution imaging by SEM and AFM is addressed, followed by Raman and surface-enhanced Raman spectroscopy. Lastly, we will discuss measurements of mass—bioaerosol time-of-flight mass spectrometry, time-of-flight secondary ion mass spectrometry, particle-induced X-ray emission/scanning transmission ion microscopy, and accelerator mass spectrometry. Results of the application of these techniques to the analysis of *Bacillus* spores will be presented, concluding with discussion and future directions.

INDIVIDUAL METHODS IN DETAIL

SCANNING ELECTRON MICROSCOPY WITH ENERGY DISPERSIVE X-RAY MICROANALYSIS

SEM is a standard "workhorse" technique for characterizing particulate samples, found in many laboratories worldwide. It provides excellent imaging of the surfaces of agent particles and other material in a sample, and is an excellent initial technique to be applied for identifying likely agent particles for analysis by other instruments. When combined with EDX, the elemental composition of the material in the imaged region can be determined.

Surface characterization and imaging of the various spore samples (e.g., *Bacillus globigii*, *Bacillus thuringiensis*, *Bacillus cereus*, and *Clostridium sporagenes*) was performed with a Hitachi S-4500 cold field emission SEM. Acceleration voltages of 6 keV and less were typically used to document spores and associated residual materials on each of the provided sample mounts. No conductive coatings (e.g., gold, platinum, or carbon) were applied to any of the samples, avoiding coating artifacts which could be misleading in characterizing the samples at higher magnifications. The spores appear to be resistant to electron beam damage; nevertheless, every effort was made to avoid beam damage to spore and residual material surfaces while imaging. This was particularly important when working with the spores in silicon aerogel matrix, as this material was easily deformed and damaged by the electron beam. Figure 13.1 shows a representative sample of SEM micrographs of varying particle morphologies which can be traced to differences in the "weaponization" processes used to generate the samples. Very little silica is observed on the *Bacillus* spores in Figure 13.1a, while Figure 13.1b shows *Bacillus* spores (indicated by the arrows) coated with fumed silica. The *Bacillus* spores in Figure 13.1c exhibit a distinct type of "shake-and-bake" silica coating in sharp contrast to the spores in Figures 13.1a and 13.1b. The *Clostridium* spore shown in Figure 13.1d shows yet another type of silica coating, i.e. colloidal silica particles.

An EDAX Genesis X-ray analysis system attached to the Hitachi S-4500 SEM was used for qualitative and quantitative elemental EDX of spore samples, including individual spores, matrix material, octahedral protein crystals (in *Bacillus thuringiensis* samples), and residual materials deposited during sample evaporation. The majority of X-ray spectra were acquired at 6 keV, allowing detection and measurement of X-ray K-lines for elements potassium and calcium, present in many of the samples. At 6 keV, L-lines for heavier metals such as chrome, iron, nickel, and copper, had they been present, would have

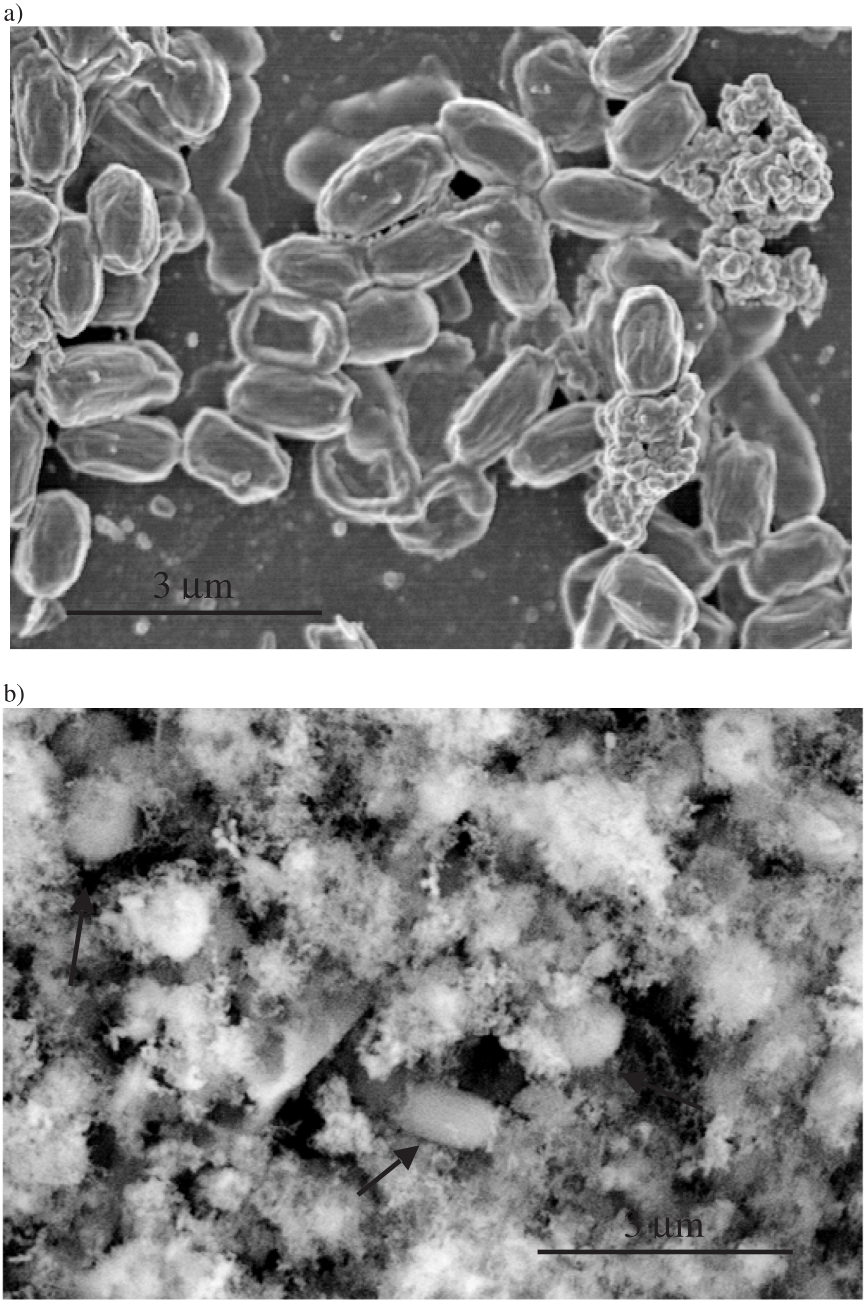


FIGURE 13.1 SEM micrographs of weaponized bacterial spores. For analysis, the powdered samples were dusted onto a sample platform of either carbon tape, silicon, substrate or TEM grid; no coatings were applied to the samples. (a) Uncoated *Bacillus* spores mixed with clumps of silica (right side of the image). (b) *Bacillus* spores, indicated by the arrows, are seen embedded in fumed silica. (c) Individual *Bacillus* spores (indicated by the arrows) coated with silica. The spores are dusted onto a TEM grid. (d) A *Clostridium* spore coated with colloidal (spherical) silica particles.

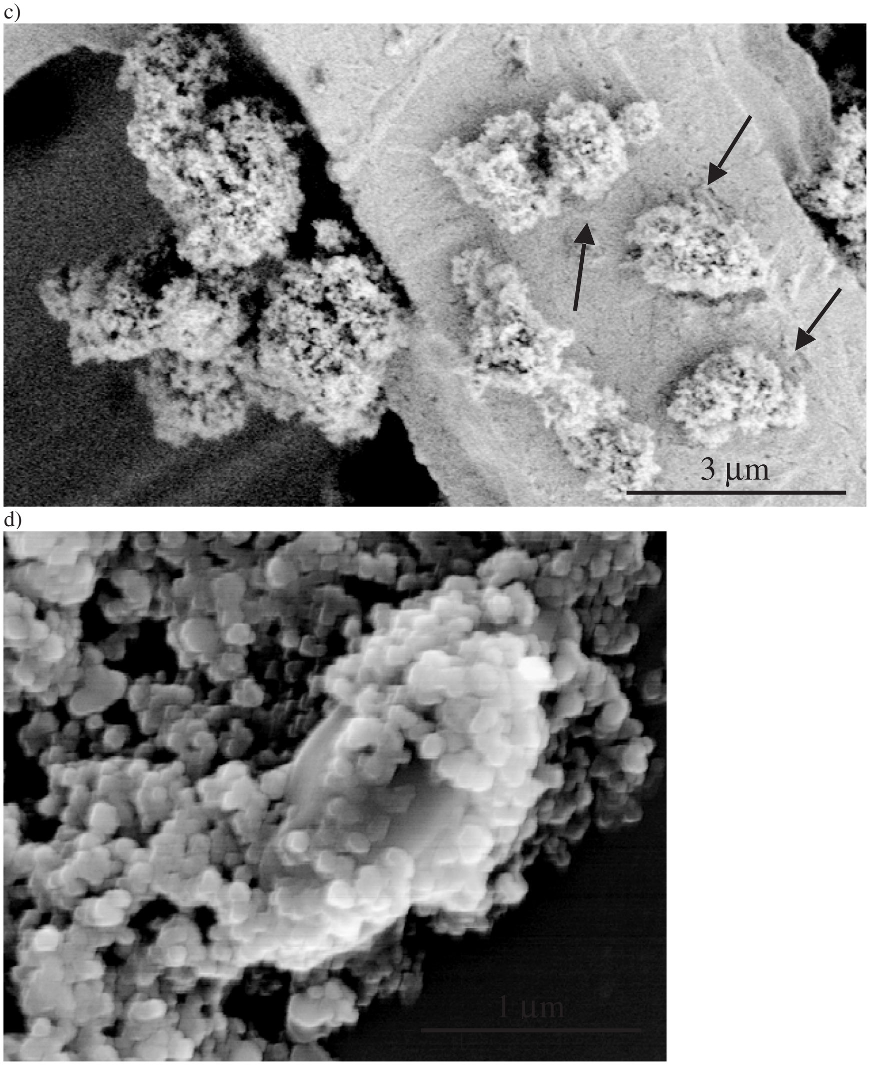


FIGURE 13.1 (*Continued*)

been measurable, along with manganese, which was detected in many of the samples. Lower-acceleration voltages were used when probing individual spores, significantly reducing the X-ray absorption and fluorescence effects of the sample mount material, primarily silicon, revealing more representative concentration count ratios for carbon, nitrogen, and oxygen in the individual spores.

Quantitative chemical analysis was performed using a standardless ZAF routine that models and subtracts background counts, calculating elemental concentrations from the remaining X-ray peaks. As the program is intended for working with infinitely thick bulk samples rather than micron-size particles, the quantitative results can only be realistically used as comparative quantities between individual spores, the matrix material, and the octahedral crystals. Precise concentrations would be difficult to determine, as electron probe depth and X-ray excitation volume at 6 keV is greater than the average spore size of 0.8 microns in diameter and 1.5 microns in length, assuming a spore density similar to that of elemental carbon.

The EDX spectra easily confirmed the presence of silica in samples where it had been added as a flow enhancer. In addition, EDX analysis helped provide the identity of certain microcrystals observed in a few samples, as shown in Figure 13.2. An SEM micrograph of a representative crystal is given in Figure 13.2a with its concomitant elemental composition shown in Figure 13.2b. Note the fairly square morphology of the crystal (~10 μm × 10 μm × 2 μm) and its high concentrations of magnesium and phosphorus. In addition, carbon, nitrogen, sodium, potassium, and calcium were also found to be present in the crystal. These microcrystal results are in sharp contrast to the octahedral protein toxin crystals generated by *B. thuringiensis*, shown for comparison in Figures 13.2c (SEM micrograph) and 13.2d (elemental composition). The crystals of bacterial origin are much smaller (~1 μm in length), and contain only carbon, nitrogen, and oxygen (the silicon is from the substrate). Since the microcrystals were observed only in samples that had been grown in a medium containing very high concentrations of magnesium and phosphate, it was surmised that they were primarily magnesium phosphate that had precipitated out of the medium during processing of the spores. In addition to this microcrystal example, other elemental signatures of various growth and processing steps including sodium, magnesium, aluminum, phosphorus, sulfur, potassium, and calcium were also identified.

ATOMIC FORCE MICROSCOPY

The technique of AFM was introduced by G. Binning[6] in 1986 and has been used to visualize biological structures for more than a decade. AFM offers several advantages when viewing biological structures. AFM measurements can be made on samples in vacuum, in air under ambient conditions, and in fluids under near physiological conditions. Samples are not required to be stained, to be conductive, or to be arranged in a highly ordered manner. AFM can be used to scan fields ranging in size from less than 20 nm up to ~150 μm,

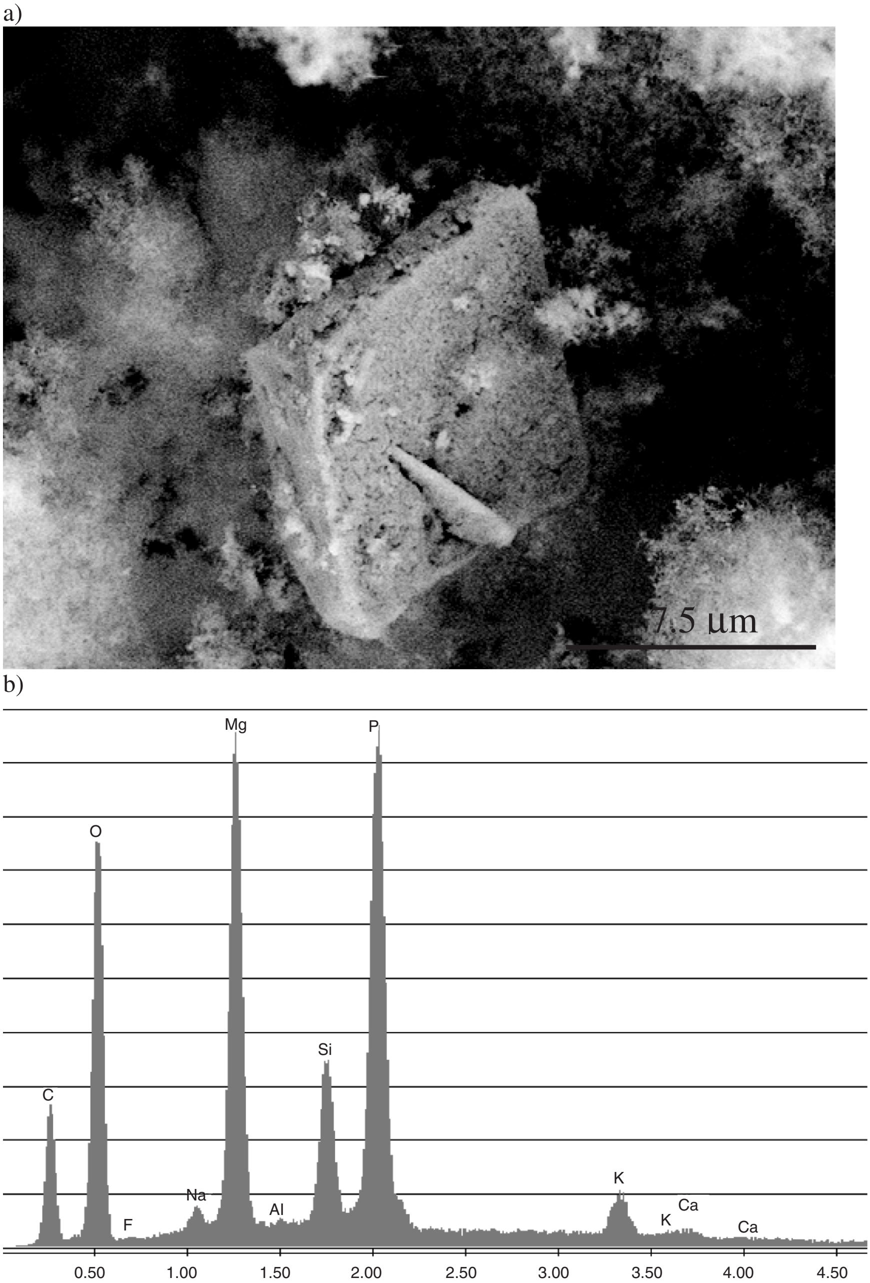
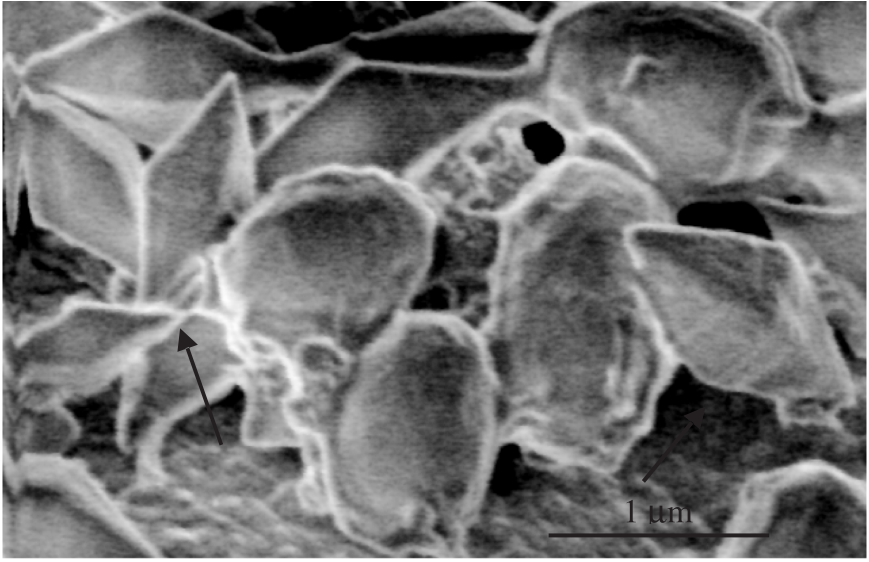


FIGURE 13.2 SEM images with concomitant EDX analysis of crystals observed in powdered weaponized bacterial spore surrogates. Powders were dusted onto a silicon substrate; no coatings were applied to the samples. (a) A micrograph of a microcrystal found in samples grown in media containing large quantities of magnesium and phosphate. The crystals are ~10 µm × 10 µm × 2 µm. (b) Elemental composition of the same crystal shown in (a), as measured by EDX analysis results in large magnesium and phosphorus peaks. (c) Octahedral protein toxin crystals (indicated by the arrows), generated by *Bacillus thuringiensis*, are ~1 µm in length. *B. thuringiensis* spores are also seen in the micrograph. (d) EDX analysis of the protein toxins seen in (c) shows the particle to be organic, containing only carbon, nitrogen, and oxygen. The silicon peak is due to the substrate.

c)



d)

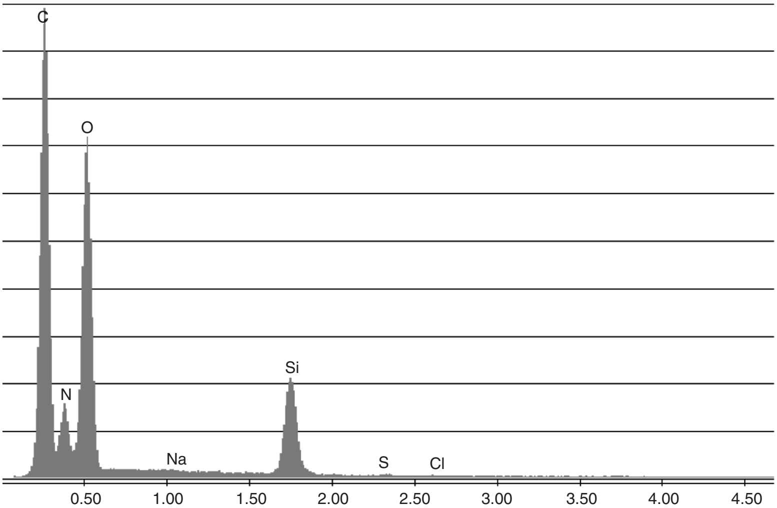


FIGURE 13.2 (*Continued*)

with a spatial resolution on biological, soft materials of ~1 nm, and a height resolution as fine as 0.1 nm. Thus it provides precise visual detail over a size range that is inaccessible for most other techniques. Its application extends over the size range lying between that of individual macromolecules (which are accessible by X-ray crystallography) to macromolecular assemblies (which

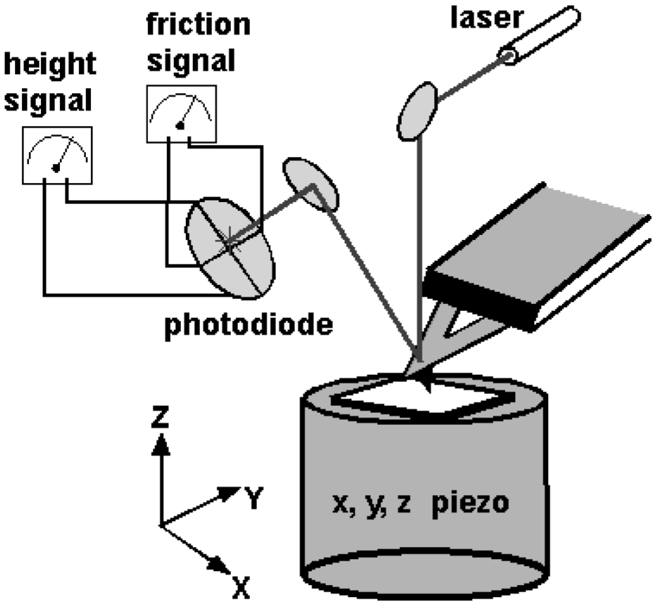


FIGURE 13.3 Schematic of an atomic force microscope.

are amenable to electron microscopy) and living cells (which can be seen using light microscopy).

Figure 13.3 is a schematic diagram of an atomic force microscope. In AFM an extremely sharp probe is brought into atomic contact with a surface and is scanned systematically over the surface. Topographic features on the surface and the tip of the probe interact with aggregate atomic forces as the probe is scanned over the surface. The sharp probe is attached to a flexible cantilever extending from a rigid substrate. A laser is reflected from the back surface of the cantilever; deflection of the tip, as it moves over surface topography, is detected as the reflection changes position on four elements of a position-sensitive diode. The differential signal between the sum signal of the top two elements and the sum signal of the bottom two elements provide a measure of the deflection of the cantilever. Similarly, the differential signal between the sum signal of the left two elements and the sum signal of the right two elements provides a measure of the torsion of the cantilever, which is a measure of the friction between the probe and sample.

When operating in "contact mode," the tip touches the sample continuously during the measurement, exerting constant normal and lateral forces as the tip is scanned over the sample. AFM applications investigating soft materials such as biological samples typically use "tapping mode" operation,[7] which minimizes contact between the probe tip and the sample surface and greatly

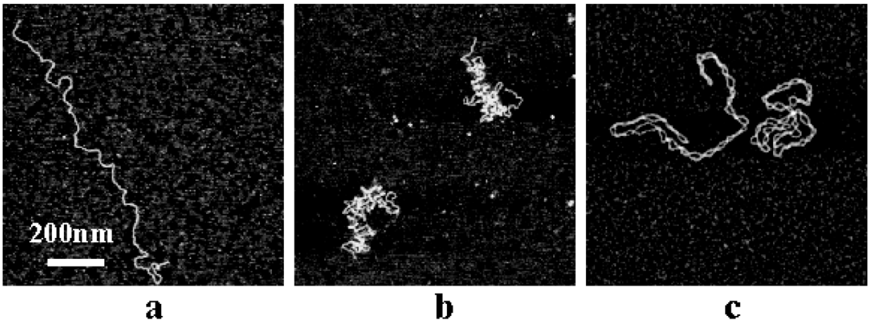


FIGURE 13.4 AFM images of DNA. (a) linear DNA, (b) linear DNA with ABFp2 packing protein, (c) highly supercoiled DNA.

reduces lateral forces. In tapping mode, the vertical position of the sample is continually adjusted by a feedback mechanism, in order to maintain constant amplitude for the freely oscillating probe. Changes in the oscillation amplitude, in response to attractive or repulsive forces between the sample and the tip, are monitored. These changes are transformed into topographic information.

The past decade has seen continuous progress in the quality of AFM imaging of biological samples. In particular, membrane proteins,[8,9] DNA,[10,11] and DNA-protein complexes[12] have been intensively studied. Examples of DNA imaged using an atomic force microscope with a standard commercial silicon tip and operating in tapping mode are shown in Figure 13.4. The image in Figure 13.4a shows linear DNA; in Figure 13.4b, ABF2p, a packing protein, was added to the DNA sample, resulting in compaction. Figure 13.4c shows highly supercoiled DNA resulting from adding TOTO-1, a dye that intercalates between base pairs.

Despite decades of study of viruses and their pressing importance in human medicine, many of their structural properties are poorly understood. Because of their heterogeneity and lack of symmetry, large viruses are often not amenable to X-ray crystallographic analysis or reconstruction by cryo-electron microscopy (cryo-EM). AFM can be effectively used to image the intact structures of large viruses,[13] and their internal structures can be revealed by AFM in combination with chemical and enzymatic dissection.[14,15] Development of AFM as a diagnostic tool to probe the structures and function of human viruses has the capacity to provide important information on their structure, function, and assembly.

Figure 13.5 shows turnip yellow mosaic virus (TYMV), a 28 nm diameter, $T = 3$ icosahedral plant virus in which the capsomeric structure of a small virus was visualized by AFM for the first time. In these images, pentameric and hexa-

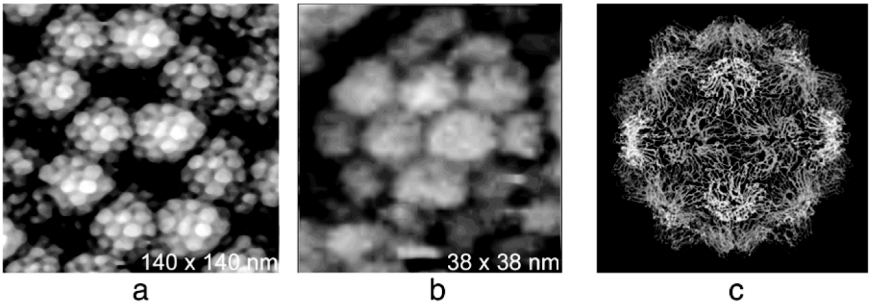


FIGURE 13.5 (a,b) *In situ* AFM images of turnip yellow mosaic virus particles immobilized in the crystalline lattice clearly display capsomers on the surface of the T = 3 icosahedral virions. (c) The structure of the capsid of turnip yellow mosaic virus based on X-ray diffraction analysis.

meric clusters which are roughly 60 Å across can be discriminated from one another, and the difference between the highest and lowest points on the capsid surface, about 45 Å,[16] was accurately reflected by AFM. Recently, the capsomeric structures and orientations of individual brome grass mosaic and cucumber mosaic virions were also visualized by AFM.[17] These studies demonstrated that AFM could provide a means for obtaining structural information directly from individual virus particles immobilized on a substrate.[17]

The clarity with which structural detail can be recorded on the surfaces of small plant viruses, such as TYMV, suggested that AFM may be even more broadly useful as an analytical tool for macromolecular structural investigations and provide important topographical information on large macromolecular ensembles such as human and other animal viruses that would otherwise be lacking. Fragility and structural heterogeneity can render such viruses troublesome targets for X-ray crystallography. The only alternative approach to date, cryo-EM, yields resolutions from 8 Å upwards.[18] A limitation of this method, however, is that it benefits greatly from high particle symmetry, such as icosahedral symmetry, and is far less powerful for irregular, polymorphic virus particles. In addition, because of a low contrast in biological EM, the high resolution requires computer averaging and processing of thousands of images.

Much of the technique used for imaging other biological structures with AFM can be applied to the imaging of bacterial spores. Figure 13.6 is an AFM image of a *Bacillus subtilis niger* spore. Figure 13.6a shows the entire spore with some structure of the protein coat visible. When a smaller area of the surface is imaged, as shown in Figure 13.6b, much more detail of the protein coat is visible. We have found that, similar to many other biological molecules, spores can be somewhat soft and sometimes exhibit regions of increased

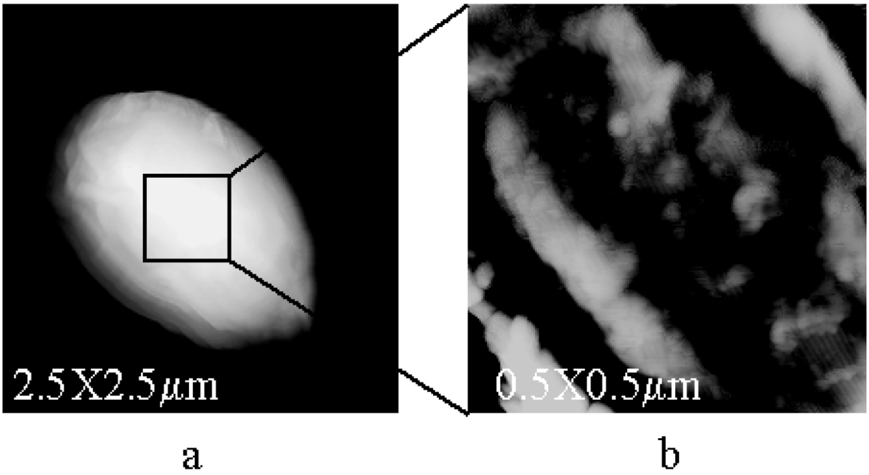


FIGURE 13.6 (a) AFM image of *Bacillus globigii* spore. (b) High-resolution image of the protein coat of the spore.

adhesion to the AFM tip. The force placed on the sample by the tip needs to be kept to a minimum so as not to deform the surface during the measurement, yet the cantilever needs to be stiff enough to overcome adhesion between the tip and sample. Imaging spores in fluid should increase resolution, since the tip does not have to penetrate the water layer that is typically on samples measured in air under ambient conditions.

Currently, the principal limitation on application of the promising technique of AFM to structural biology is the resolution limit imposed by the finite tip radius—which is typically larger than 10 nm. Overall tip sharpness becomes more important in imaging the relatively rough structures of pathogens having deep crevasses, and in delineating details of protein complexes. Recently, carbon nanotubes (CNTs) have emerged as the next generation of force microscopy probes.[19–22] These probes typically have an aspect ratio of more than 100, and the end radius of curvature of a single walled nanotube is ~10 Å. This small tip radius, combined with the high aspect ratio and mechanical robustness, presents an obvious advantage over conventional AFM probes and permits resolution about one order of magnitude better than that achievable with commercial AFM tips. High-resolution imaging capabilities of CNTs were demonstrated recently on various biomolecules, such as DNA, antibodies, proteins, and nucleosomes.[21]

RAMAN SPECTROSCOPY AND SURFACE-ENHANCED RAMAN SPECTROSCOPY

Raman scattering is the inelastic scattering of photons off of molecular bonds. The Raman scattered photons differ in frequency from the incident and elastically scattered photons by the frequency of the molecular bond vibration. Figure 13.7 shows an energy level diagram illustrating the Raman scattering process. Most of the incident light is elastically scattered (Rayleigh scattered); however, a small fraction of the incident light (~1 in 10^8 photons) is Raman scattered. Although the Raman scattering effect was first described in 1928, it did not find wide application until the development of lasers in the 1960s.[23–25] The laser provides an intense, monochromatic light source that is ideal for

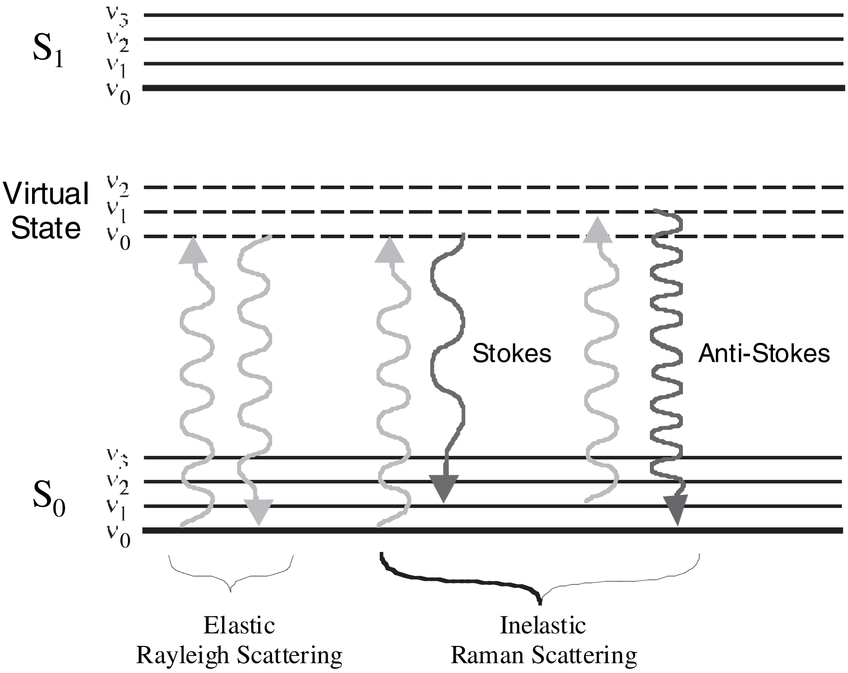


FIGURE 13.7 Energy level diagram illustrating the Raman scattering process. Most of the light incident on a chemical bond is elastically scattered, resulting in no change in the frequency of the photon. However, a small fraction of photons are inelastically, or Raman scattered off the molecular bond vibration. In this case the molecule either returns to a higher vibrational energy level, yielding a Stokes shifted photon, or it returns to a lower vibrational energy level, leaving an Anti-Stokes shifted photon.

Raman spectroscopy. Since the development of lasers, Raman spectroscopy has become an important analytical tool for a wide range of disciplines.

Raman spectroscopy provides a wealth of information that can be used to both identify and quantify the molecular species under study. Chemical bonds have unique vibrational frequencies; therefore different molecules have distinct Raman spectra. By comparing the spectroscopic peak positions and relative intensities to those available in a vast library of Raman spectra, molecules can be identified and quantified. Moreover, the information acquired with Raman spectroscopy is complimentary to infrared (IR) absorption spectroscopy, which also measures the frequency of bond vibrations. This provides an additional database for comparing Raman spectra and identifying a particular Raman transition.

Although Raman spectroscopy has gained wide acceptance for analytical chemical applications, its use in biological studies has been somewhat more limited. One reason for the lack of application to biology is the complexity of the signals acquired from biological materials. Biological molecules are typically large, and since the number of vibrational modes is $3n - 6$, where n is the number of atoms in the molecule, the spectra can become too congested to interpret. Another reason for the limited application of Raman spectroscopy to biological studies is the intrinsic autofluorescence that can overwhelm the Raman signal. While the Raman spectra for biological samples remain complex, recent advances in near-IR detectors have allowed Raman spectra to be collected with near-IR lasers, thereby reducing the autofluorescence background. As a result, Raman spectroscopy is becoming a more common tool for investigating biological systems.

Raman spectroscopy is particularly useful for identifying conformational states of biological molecules.[26,27] It has been applied to study the conformation of lipids in cell membranes, DNA, and proteins. Studies on proteins, for example, have revealed that the amide stretching frequencies in the polypeptide backbone can be used to identify the secondary structure present in the molecule.[26,27] In the Raman spectra of proteins, typically only the amide I and amide III bands are observable. The amide I peak is around $1650\,\mathrm{cm}^{-1}$ for α-helices and shifts to near $1675\,\mathrm{cm}^{-1}$ for β-sheet conformations. Similarly, the amide III peak ranges from $1225\,\mathrm{cm}^{-1}$ to $1245\,\mathrm{cm}^{-1}$ depending on the secondary structure of the proteins. Additionally, other spectroscopic signatures, such as S-S vibrations present in cysteine cross-linking as well as the peaks from aromatic amino acids, help to determine the secondary structure and identity of the protein.

As the library of spectroscopic signatures of biological molecules becomes larger, Raman spectroscopy is beginning to find applications in the biomedical field. Early cancer detection as well as the detection of precancerous cellular changes, the characterization of atherosclerotic plaques, and the identification of pathogenic organisms have all been identified as potential clinical applications for Raman spectroscopy.[28,29] Along these lines, bacteria and their endospores have been studied using Raman spectroscopy.[30] Early work applying Raman spectroscopy to lyophilized bacterial spores revealed that the primary spectral features observed were due to the calcium dipicolinate present in the cortex of the spore.[31] Later work using resonance-Raman spectroscopy to study bacterial spores showed significant differences between the Raman spectra of dormant spores and vegetative bacterial cells.[32] This difference was attributed to the release of calcium dipicolinate upon germination. Although other spectral features are observed, such as the amide bands corresponding to the secondary structure of the proteins incorporated in the spore, as well as transitions from some aromatic amino acids, these structural features do not change significantly upon spore germination.

In the present study, we extended the characterization of bacterial spores to the single-spore level. A customized micro-Raman setup (see Fig. 13.8a for a description) allows the sample to be scanned and Raman spectra acquired with spatial resolution in the xy plane on the order of 1 µm. Samples are then diluted and prepared on a solid substrate such that individual spores are spatially isolated. Figure 13.8b shows a 20 µm × 20 µm area of a sample containing *Bacillus* spores dried onto a calcium fluoride substrate. The bright spots in the image are the fluorescence from individual *Bacillus* spores. In order to collect the Raman spectra, the spore is centered on the focused laser beam and photobleached for approximately one minute. Once the background fluorescence is reduced in this fashion, the Raman spectra can be collected in as little as 20 seconds.[33] The Raman spectra for four different species of *Bacillus* spores are shown in Figure 13.9. What is striking about the spectra shown in Figure 13.9 is the similarity between the different spore species. The nearly identical spectra are the result of the contribution from calcium dipicolinate. Calcium dipicolinate is present in the cortex of the spore and comprises ~10% of the dry weight of the spore. While other spectral features that could possibly provide a means of distinguishing between spore species are also clearly present, they are overshadowed by the strong calcium dipicolinate signal.

In order to circumvent this problem, we have employed SERS. SERS is a process through which the Raman signal is greatly enhanced when the chemical species of interest is chemisorbed to a metal (typically gold or silver) surface.[34] When the metal surface is illuminated with the appropriate laser wavelength, a surface plasmon resonance is excited. When the surface plasmons encounter sharp discontinuities in the metal surface, the electric field leaks from the surface at the point of the discontinuity and decays exponentially with distance from the surface, similar to evanescent waves in total internal reflection spectroscopy. This has the effect of focusing the electric field and

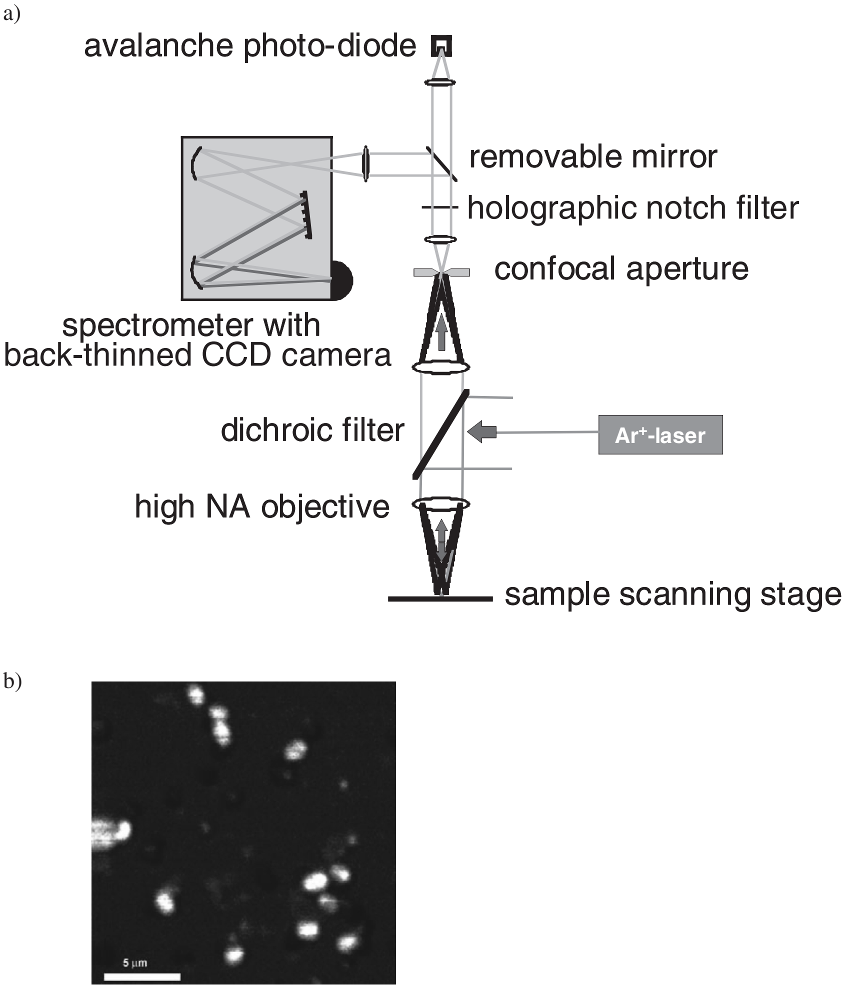

a)
avalanche photo-diode
removable mirror
holographic notch filter
confocal aperture
spectrometer with back-thinned CCD camera
dichroic filter
Ar+-laser
high NA objective
sample scanning stage
b)
5 µm

FIGURE 13.8 Schematic illustration of confocal microscope used for the micro-Raman experiments. (a) The 488-nm line of an Ar^+ laser is focused onto the sample with a high numerical aperture (NA) objective. The sample is mounted on an xy-piezo stage and raster-scanned over the focused laser beam. Raman scatter is collected with the same high-NA objective and focused onto a confocal aperture. The Rayleigh scatter is removed with a holographic notch filter, and the remaining Raman scatter is focused onto an avalanche photo-diode, the signal which is used to build up the image. Once a feature of interest is located, the laser beam is centered on the object and the Raman scatter is sent into a spectrometer with a back-thinned CCD camera to acquire the Raman spectra. (b) Confocal image of individual bacterial spores dried onto a calcium fluoride substrate. The bright spots correspond to the intrinsic autofluorescence and Raman scatter from the spores. Once the autofluorescence is photobleached, the Raman spectrum from the individual spore is collected. (See color insert.)

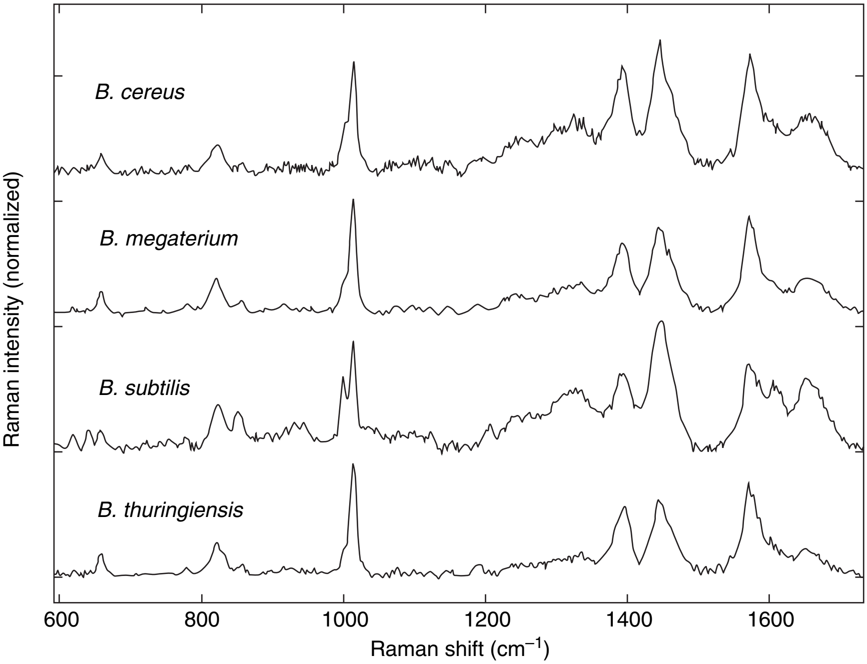


FIGURE 13.9 Micro-Raman spectra of individual spores from four different *Bacillus* species. Each spectra was collected using the 488-nm line from an Ar$^+$ laser. The signal was integrated for 4 minutes for each spectra. The nearly identical spectra can be attributed to the intense signal arising from calcium dipicolinate in the cortex of the spore.

therefore providing enhanced Raman signals. Enhancements to the Raman signal for species adsorbed onto roughened metal surfaces have been reported to be as high as 10^6.[34] Recently, however, it has been demonstrated that by using metal colloids (50–100 nm in diameter) as the SERS substrate, the enhancement factors can be as large as 10^{15}.[34] This enormous enhancement to the Raman signal has made it possible to acquire Raman spectra for single molecules.[35,36]

Using metal colloids as the SERS substrate also has an additional advantage for biological applications. Because the SERS effect only enhances the signal a few nanometers from the metal surface, the metal nanoparticles act as near-field probes of the structures to which they are attached. This provides a means to probe the local, nanometer-scale, structure, and chemical composition, which would not be possible with conventional Raman spectroscopy. An illustration of using SERS particles as local probes is demonstrated again with *Bacillus* spores. In this sample the *Bacillus* spores have been incubated with colloidal silver suspension. The silver colloid is chemisorbed to the outside of

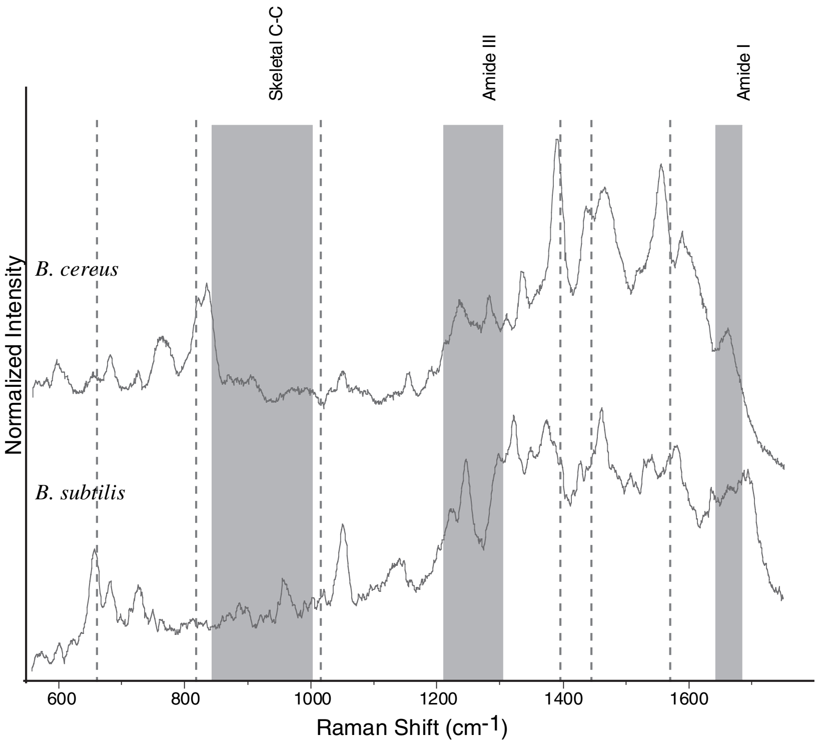


FIGURE 13.10 Surface-enhanced Raman spectra from colloidal silver particles attached to the outside of the spores. The dashed lines show where calcium-dipicolinate peaks would be if they were present in the spectra. Also highlighted are the regions used to identify the secondary structure of proteins.

the spore. Figure 13.10 shows the spectra collected from silver nanoparticles attached to spores of two different *Bacillus* species. What is evident from the spectra is that they are no longer dominated by the calcium dipicolinate as seen with the conventional Raman spectra in Figure 13.9. This result is consistent with the location of the calcium dipicolinate (DPA) in the cortex of the spore, several hundred nanometers away from the nanoparticle probe, thus providing a means to probe the outside structure of the spore.

In summary, we have shown that micro-Raman spectroscopy and surface-enhanced Raman spectroscopy provide chemical information of biological materials on the sub-micron scale. These are important tools for forensics analyses because of their high spatial resolution, good chemical selectivity, and high sensitivity. We have demonstrated this by presenting spectroscopic results obtained from single, isolated bacterial spores.

BIOAEROSOL TIME-OF-FLIGHT MASS SPECTROMETRY

BAMS is a novel real-time technique for the rapid identification of individual bioaerosol particles using mass spectrometry. BAMS is unique in that it allows for the reagentless analysis of a complex environmental sample without any prior sample preparation or discrimination. In effect, it offers true real-time (i.e., instantaneous) results from bioaerosols sampled directly from the atmosphere within a large background of both biological and nonbiological particles without the sample preparation required in PCR and other molecular-based assays. *Bacillus* spores have been successfully characterized and can be efficiently distinguished from other *Bacillus* spores at the species level, fungal spores, vegetative bacteria, and many other biological and nonbiological background materials in real-time. Indications of the metabolic processes that lead to the creation of *Bacillus* spores have also been observed.

BAMS analyzes individual aerosol particles by single-particle laser desorption/ionization time-of-flight mass spectrometry. Many incarnations of this technology have been demonstrated.[37] We chose aerosol time-of-flight mass spectrometry because of its ability to accurately size individual particles, its use of a desorption/ionization laser at 266 nm, and its ability to collect both positive and negative mass spectra simultaneously from the same particle.[38,39] The 266-nm laser wavelength is fortuitous because it corresponds to the absorbance maximum of dipicolinic acid, a component of *Bacillus* spores that constitutes up to 15% of their dry weight.

In aerosol time-of-flight mass spectrometry, many particles are introduced into the instrument through a converging nozzle where each is accelerated to a terminal velocity as a function of its aerodynamic diameter. A particle continues through differentially pumped stages until it reaches a region of ultrahigh vacuum. There, it encounters two lasers, spaced a fixed distance apart, that intersect its path. It scatters light from each laser to a corresponding photomultiplier tube. The time difference between the light scatterings indicates the particle's velocity, which corresponds to its aerodynamic diameter and also predicts its time of arrival at the center of the ion source region of a bipolar time-of-flight mass spectrometer. The particle is desorbed and ionized by a Nd:YAG laser operating at 266 nm, and its chemical composition is measured by the mass spectrometer in both polarities. The spectra are reported to a data analysis computer for classification. The data analysis takes place on a separate computer that is capable of monitoring an arbitrary number of mass spectrometers remotely. The algorithm for the analysis of the spectra is similar to the ART-2a and fuzzy c-means algorithm that have been previously used with data from single-particle mass spectrometers.[40–43] As the particles are identified, their identities are plotted in a pie chart, and select threat agents trigger a visual and optional audible alarm.

Our objective was to demonstrate a technology that could operate autonomously, consume only electricity, and provide species-level detection of *Bacillus* spores within a background of both biological and nonbiological aerosols. To date we have analyzed more than 20 different growth media, and a thorough study of them is forthcoming; however, LB agar was analyzed to represent growth media for the purposes of these experiments. Simple biological molecules were represented by aspartame in Equal Sweetener, while collagen from Knox Unflavored Gelatin represented purified biological extracts. Mineral aerosols were represented by Gold Bond medicated powder, which contains zinc oxide. The BAMS system was originally deployed to detect *Bacillus anthracis* in the U.S. mail, and so the above challenge materials were selected because they are widely available white powders that, to the casual observer, might appear to be *Bacillus* spores. In addition, several bacterial and fungal spores were used to demonstrate specificity of the technique.

To demonstrate that our technique could identify the endospores within a complex background, a series of experiments was performed where bacterial spores were nebulized in an aerosol of the other materials. Figure 13.11 shows the results of these experiments, the spectra categorized by the material detected. All three samples were prepared separately, and thus the concentrations of the background aerosols differ between them. Obviously, however, BAMS can detect bacterial spores in a complex mixture. Figure 13.11a shows a mixture of the nonbiological test samples. The large "Other" category in Figure 13.11a results from the large amount of gelatin in the mixture. Gelatin is categorized as "Other" more often than it is recognized, and in Figure 13.11a, 20% of all spectra were recognized as "Gelatin" and 29% as "Other."

Figure 13.11b shows the same mixture with *Bacillus* spores added, along with baking soda and powdered sugar, two previously uncharacterized samples. Note that a large fraction of the particles are called "Other" in response to the addition of the unknown samples. Figure 13.11c shows the mixture with *Clostridium* cells nebulized into the sample. They were not detected in Figure 13.11b, and fungal spores, which are still absent, are not detected in Figure 13.11c. Although the *Bacillus* spores run in the tests shown were from *B. globigii*, separate field tests run at an offsite biosafety level 3 facility have demonstrated the ability of BAMS to recognize *B. anthracis* from within a complex mixture as well.

BAMS may ultimately prove a very valuable tool for the interrogation of biological systems at the level of their most fundamental elements. As currently implemented, it has already proven itself to be a robust method of detecting individual *Bacillus*, *Clostridium*, and fungal spores and cells against a variety of biological and nonbiological background materials. Although the applica-

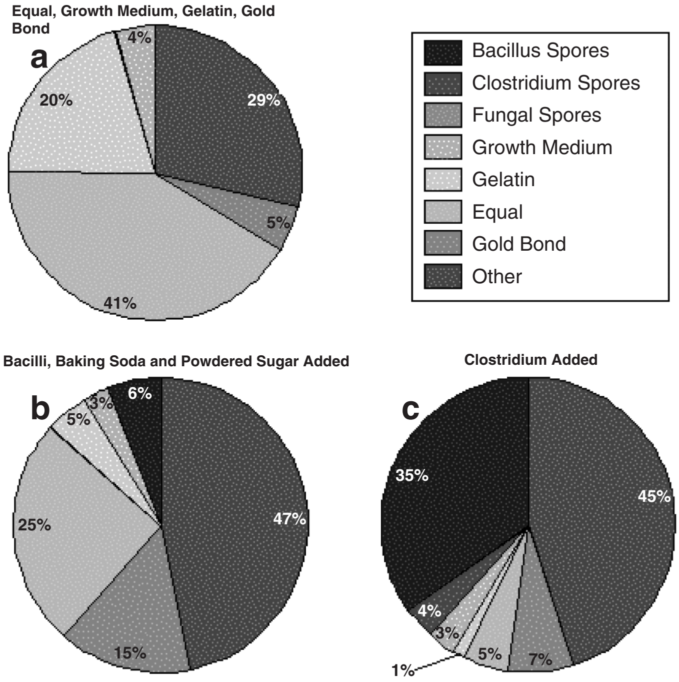


FIGURE 13.11 BAMS at work: Three samples of particles analyzed by our real-time spectrum identification software. (a) Only the background materials have been added. (b) In addition to the background materials, *Bacillus* spores have been added. (c) In addition to the background materials and *Bacillus* spores, *Clostridium* spores have been added. (See color insert.)

tions presented in this study deal with the detection of biological attack, with future improvements as a single-cell analysis technique, BAMS will also have far-reaching contributions in the fields of oncology, microbiology, and public health.

TIME-OF-FLIGHT SECONDARY ION MASS SPECTROMETRY

ToF-SIMS is a microanalysis technique that characterizes the composition and distribution of the sample surface constituents.[44–46] ToF-SIMS involves bombarding the sample surface with a pulsed ion beam (primary ions). As the result of ion impact, the species from the surface are ejected, ionized (secondary ions), and accelerated into a mass spectrometer. These secondary ions are mass-analyzed by measuring their time-of-flight from the sample surface to the detector. Collectively, the ToF-SIMS technique provides spectroscopy for characterization of chemical composition, imaging for the determination of the distribution of chemical species, and depth profiling for thin-film characterization.

Categorized as a static SIMS technique, ToF-SIMS analysis involves the application of an ion beam of intensity $<10^{13}$ primary ions/cm^2 (or defined as sampling equal to or less than 1% of a monolayer) which analyzes the molecules at the top layer of the sample surface. Therefore, the technique is highly surface-sensitive and provides information from the outermost layer of the sample surface. Additionally, static SIMS maximizes the quality and quantity of molecular information obtained by significantly reducing fragmentation of sputtered ions.

By rastering a finely focused ion beam across the sample surface, ToF-SIMS imaging analysis provides a map of surface chemical and elemental distribution. Due to the parallel detection nature of ToF-SIMS, the entire mass spectrum is acquired from every pixel in the image. The mass spectrum and the molecule-specific images can then be reconstructed to determine the composition and distribution of sample surface constituents. The ultimate image spatial resolution is limited by the spot size of the ion beam at ~150 nm in diameter.

A schematic of the ToF-SIMS instrument is given in Figure 13.12. ToF-SIMS analyses were performed on a Physical Electronics TRIFT III ToF-SIMS instrument (Physical Electronics, Eden Prairie, MN, USA), equipped with three primary ion sources: gallium ($^{69}Ga^+$) liquid metal ion gun (LIMG), dual plasmatron oxygen (O_2^+) gun and a cesium (Cs^+) ion gun.[47] Results presented here were obtained using gallium ions as the primary ion source. The LIMG can be operated at either 25 keV, 600pA mode with a spot size of <150 nm for imaging, or in 15 keV, 3nA mode to achieve high mass resolution ($m/\Delta m \sim 8{,}000$ at $m/z = 41$). A low-energy pulsed electron beam, fired between two ion beam pulses, can be used for charge compensation on the nonconductive samples.

Although widely accepted as a high-sensitivity analytical technique for characterization of surface organic and inorganic compositions, the application of ToF-SIMS to biological materials is still in its relatively early stages. The typical examples are in the areas of characterization of biomolecules on surfaces such as microarrays.[48,49] Due to the intrinsic limitation on the working mass range of the technique, the ToF-SIMS has been used primarily to study small molecules. However, efforts have been made to increase the ionization yields and thereby extend the detectable mass range for ToF-SIMS.[50,51] Alternatively, using a multivariate statistical approach such as principal component analysis (PCA), several groups are focusing their studies on the identification

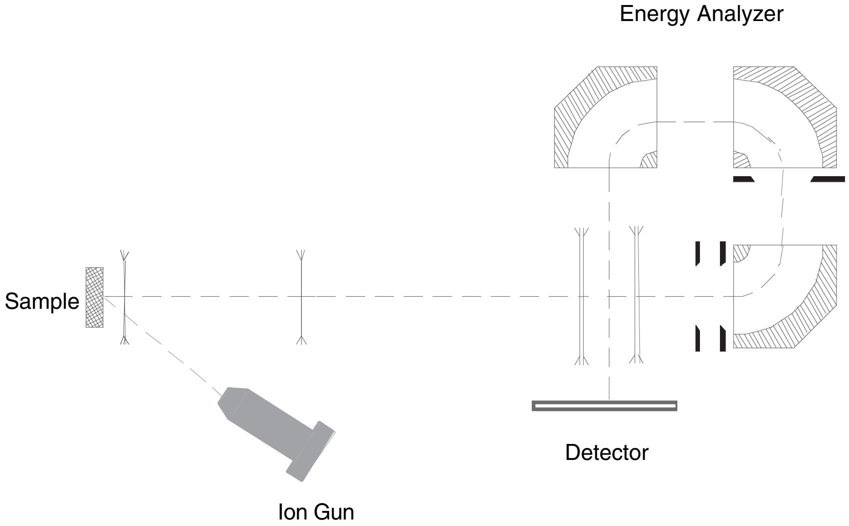


FIGURE 13.12 Schematic of the time-of-flight secondary ion mass spectrometry instrument.

of proteins and compounds from complex biological samples.[52,53] PCA is a pattern recognition technique that improves the ability to analyze the complex ToF-SIMS data set. However, such a technique provides no specific spatial information.

Molecular imaging by static SIMS has gained considerable interest in the past few years.[54,55] The key advantage in this area is the capability of direct identification of molecular components in the biological samples. The most notable example is the identification of cellular sections following freeze fracture, in which extensive chemical information of the fracture surface and cellular sections were determined.[56,57]

NUCLEAR MICROSCOPY (PIXE/STIM)

The proton or (nuclear) microprobe quantitatively maps simultaneous element distributions within microscopic regions of a sample via the use of particle (proton)-induced X-ray emission (PIXE), mass by scanning transmission ion microscopy (STIM) and low Z elements ($Z < 12$) through nuclear reactions or scattering. This technique provides spatial resolution down to 1 μm in a sample, and has been used to successfully perform quantitative analyses of element distributions in individual cells and thin-tissue sections for toxicology, physiology, structural biology, and biochemistry applications. It has also been used to analyze native electrophoretic gel blots for bound ligands, with same sample analysis of protein mass for stoichiometric relationships.[58–62] For our application to bioweapon surrogate spore samples, both bulk and individual particulate analysis was performed.

The microprobe source, accelerator and beam lines occupy an area of 3 × 20 meters. Starting at one end, $H^{\vee}$ or O^{-} ions are produced using a gas-fed ion source. The $H^{\vee}$ or O^{-} ions are then mass/energy-analyzed by the low-energy injection magnet and injected into a National Electrostatics Corporation 5SDH-2 tandem accelerator. The maximum operating voltage of the accelerator is 1.7 MeV, but normal operating voltage is 1.5 MeV, to produce ~3.0 MeV protons or 6.0 MeV O^{3+}. The accelerated ions are collimated and focused on samples contained in the sample chamber at the end of the beam line.

With PIXE, a beam of MeV energy protons is used to eject inner shell electrons from specimen atoms. When outer shell electrons fill the resulting vacancies, characteristic X-rays whose energies identify the particular type of atom are emitted. X-rays from multiple elements ($Z > 11$) are simultaneously detected. PIXE quantifies elements ($Z >$ Na) by counting elementally characteristic X-rays emitted from a defined spatial region when the sample is struck by a known amount of accelerated protons. PIXE is analytically quantitative with elemental sensitivities that can approach 0.1 mg/kg.[63]

Microbeam PIXE (μ-PIXE) utilizes a focused beam that is scanned across the sample. X-rays from multiple elements are simultaneously detected and recorded as a function of position as the beam is rastered across the sample. The proton beam can be focused to under 1 μm in diameter for investigating detailed element distributions in single cells, isolated cellular components, or tissue slices. μ-PIXE analysis of thin-film standards with biological samples quantifies elemental abundance to greater than 95% accuracy. By selecting X-rays of a specific energy, maps of element distributions can be obtained for the collected data set. Figure 13.13 illustrates this capability and shows the zinc distribution (attograms) in an 8 μm by 8 μm scan of a red blood cell.

STIM also utilizes a focused beam that is scanned across the sample. The energy loss is detected as a function of position as the accelerated ions (H^{+} or O^{3+}) pass through the sample. This enables the mass of the sample to be accurately quantified without specific standards. We have demonstrated this on the analysis of isolated proteins with the quantification of 50 ng of bovine serum albumin. Figure 13.14 demonstrates the combination of PIXE to measure elemental quantities and STIM to measure protein mass. In Figure 13.14, we see the location of ferritin in a nitrocellulose membrane from a blot of an electrophoretic gel. The figure is oriented such that the gel's well is on the left. The top portion of the figure shows the location of several elements (phosphorus, sulfur, calcium, and iron) in the gel, while the bottom graph indicates the co-location of proteins and iron, relative to the starting well. The presence

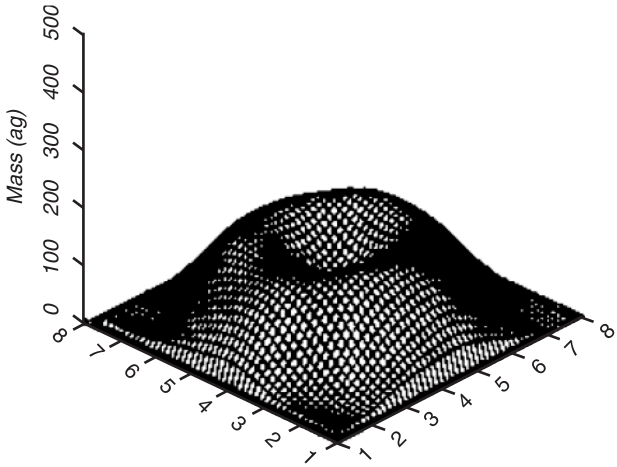


FIGURE 13.13 Zinc distribution in a red blood cell. Quantities are measured in attograms (10^{-18} gm). Note the even distribution throughout the cell.

of iron, as shown by the red line, is located at two positions; proteins are indicated by the blue line and are measured in ng (10^{-9} gm). Ferritin, an iron-binding protein, is shown at ~0.8 cm from the well, the overlap of the protein and iron curves.

ACCELERATOR MASS SPECTROMETRY

The Center for Accelerator Mass Spectrometry (CAMS) operates a dedicated AMS system that is housed in an 8,000 square-foot facility. Currently, this system performs ~25,000 analyses per year for isotopes of 10 different elements, which makes it the most versatile and productive AMS system in the world. This system is particularly well suited for the precise analysis of carbon-14 (^{14}C) in organic samples, which makes it amenable for the precise measurement of this isotope in bacterial samples.

AMS uses mass spectrometry and nuclear detection to directly measure the concentration of an isotope in a sample, eliminating inefficiencies from decay counting of radioisotopes. AMS uses two fundamental “tricks” to resolve isobars: molecular isobars dissociate by collisional charge-changing from negative ions to multiply-charged positive ions; and nuclear isobars are distinguished in particle detectors after mass and energy selection. The “accelerator” is required for AMS because the molecules to be broken, mainly hydrides and

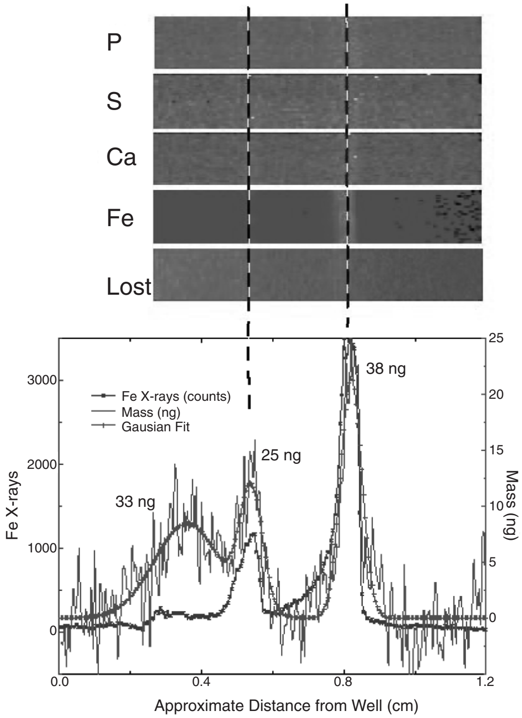


FIGURE 13.14 Location of ferritin in an electrophoretic gel. The figure is oriented such that the gel's well is on the left. The top portion of the figure shows the location of several elements in the gel as identified by PIXE analysis. The bottom graph indicates the co-location of proteins and iron in the gel. Iron was measured by PIXE simultaneously with STIM measurements of protein mass. Ferritin, an iron-binding protein, is shown at ~0.8 cm from the well, the overlap of the protein and iron curves. (See color insert.)

other binaries, are only broken at MeV energies, rather than the keV energies normally used in standard isotope-ratio mass spectrometry.

In the 10 MeV AMS system in use at CAMS, small samples of biological origin are introduced to the spectrometer after being combusted to carbon dioxide and reduced to graphite. A schematic of the system is shown in Figure 13.15. Carbon atoms and molecular isobars are subsequently ionized and accelerated toward a positive potential of 6.5 MeV in a tandem Van der Graff accelerator. These 7 MeV negative ions collide with a thin foil within the accelerator to lose five electrons and become positively charged. Molecular ions (i.e., ^{12,13}CH) do not survive at this charge state and are thus eliminated as interfering isobars. The 4^+ ion then obtains an additional 4 × 7 MeV in accelerating away from the high-potential collision cell to become a 35 MeV $^{14}C^{4+}$

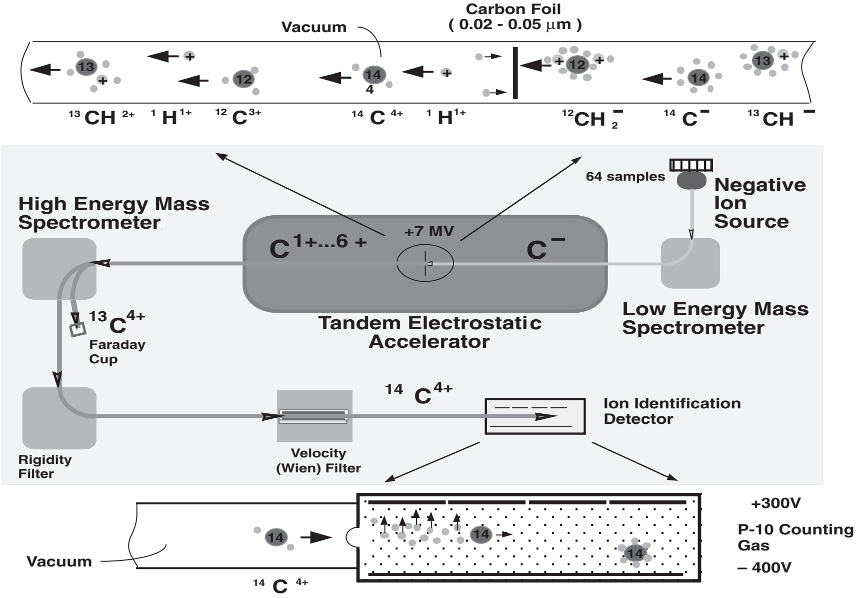

Carbon Foil
(0.02 - 0.05 µm)
Vacuum
13
12
14
4
12
14
13
13 CH 2+
1 H1+
12 C3+
14 C 4+
1 H1+
12CH 2 −
14 C −
13 CH −
64 samples
Negative Ion Source
High Energy Mass Spectrometer
C 1+...6 +
+7 MV
C −
Tandem Electrostatic Accelerator
Low Energy Mass Spectrometer
13 C4+
Faraday Cup
14 C4+
Ion Identification Detector
Rigidity Filter
Velocity (Wien) Filter
+300V
P-10 Counting Gas
− 400V
Vacuum
14
14
14
14 C 4+

FIGURE 13.15 Schematic of the 10 MeV accelerator mass spectrometry system. (See color insert.)

ion. Such ions are readily resolved by high-energy mass spectrometry and are subsequently identified by characteristic energy loss mechanisms in a gas-ion detection cell. Using this approach, AMS can be used to quantify attomoles (i.e., 10^{-18} moles) of ^{14}C in samples of size <1 mg at a precision of better than 1%. No other analytical method is capable of achieving this sensitivity or precision for ^{14}C.

All living things accumulate ^{14}C from their nutrients and food and maintain a ^{14}C concentration that reflects their average intake. At the cessation of nutrient intake, which occurs at death or during periods of dormancy, the ^{14}C content of the organism decreases exponentially with time. This is the basis for traditional carbon dating, which is routinely used to determine the age of organic materials. However, because relatively large amounts of ^{14}C were created during atmospheric nuclear tests in the 1950s and early 1960s, all living things have elevated levels of ^{14}C that can be used to estimate their age since 1950. This measurement does not depend on the radioactive decay of ^{14}C, but rather on the elevated concentrations of atmospheric ^{14}C that have been accumulated by all living things.[64–67]

Recent AMS analyses at CAMS have demonstrated that precise ^{14}C measurements can be determined from relatively small (<1 mg) samples of bacterial cultures. These cultures were from archived collections dating back to the 1950s. They clearly showed the incorporation of "bomb ^{14}C" and yielded radiocarbon contents that were consistent with the age of their production.

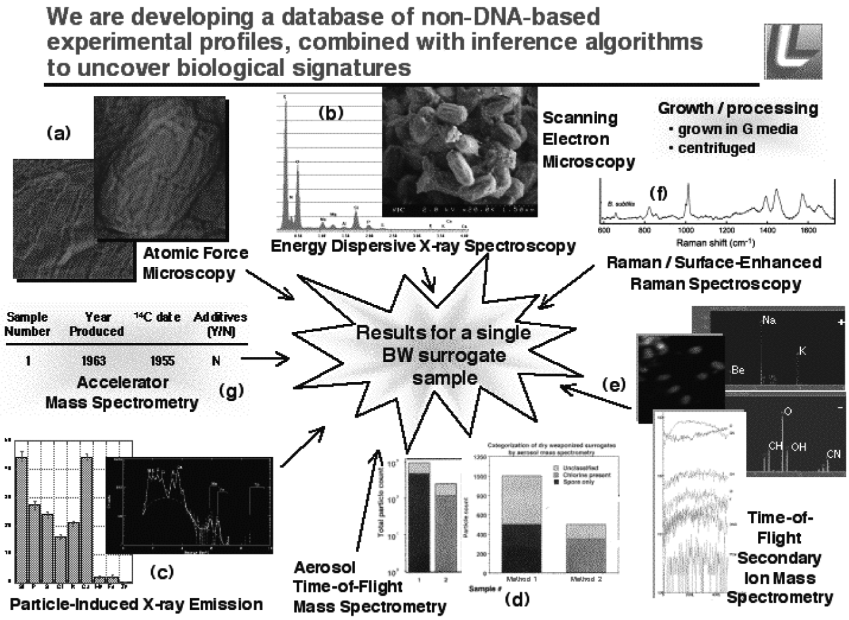
In addition to the possibility of identifying culture dates, isotopic ^{14}C signatures may make it possible to discern whether bacteria in question are not from the same culture. For example, bacteria grown on the same carbon media will contain identical ^{14}C concentrations. However, bacteria grown on media that were produced at different times would have their own distinctive ^{14}C content. With this approach it may be possible to discriminate the sources of otherwise genetically identical organisms by their ^{14}C signatures.

RESULTS: ANALYSIS OF *BACILLUS GLOBIGII* SPORE SAMPLES

The AFM results shown in Figure 13.16a are from *Bacillus globigii* spores which we grew ourselves, and were not processed in any way beyond being washed three times with distilled water. The image on the right in Figure 13.16a is a low-resolution ($1.6\,\mu m \times 1.6\,\mu m$) AFM image of a single *B. globigii* spore dried onto a mica substrate. Note that to its left, at high resolution ($450\,nm \times 450\,nm$), we are able to determine the presence of rods of proteins, with a periodicity of ~8 nm. Our findings of this structure are the first to be reported for *Bacillus* spores,[68] but the structures have been observed and termed "rodlets" on fungal spores.[69,70] Although AFM is a relatively new technique in forensic investigation, it promises to provide important information on molecules adhering to spores and on modification to the exosporium caused by mechanical and chemical treatments during weaponization processes.

We have also applied these methods to several samples of weaponized surrogate *Bacillus* spores from external sources and have included here the results of our analyses of one of those weaponized samples. A summary of the results can be seen in Figures 13.16b–g. A cursory characterization indicates that the sample was in powdered form, appeared tan in color, and was composed of fine particles; the spores were *Bacillus globigii* which had originally been grown in G media. Their viability was determined using a standard method of dilution to be 2.1×10^{11} CFU/gm. Heat shocking the spores prior to growth resulted in an approximate one order-of-magnitude increase in measured viability.

Our initial step in the analysis of the powdered surrogate was to image the sample with SEM and analyze the elemental composition with the attached EDX instrument. For our analysis, we dusted the powder onto carbon tape but found it unnecessary to coat it with an electrically conducting material; this approach allowed much more structural detail of the bulk sample to be


We are developing a database of non-DNA-based experimental profiles, combined with inference algorithms to uncover biological signatures
(a)
Atomic Force Microscopy
(b)
Energy Dispersive X-ray Spectroscopy
Scanning Electron Microscopy
Growth / processing
• grown in G media
• centrifuged
(f)
Raman shift (cm⁻¹)
Raman / Surface-Enhanced Raman Spectroscopy
Results for a single BW surrogate sample
(e)
Na
K
Be
O
CH
OH
CN
Time-of-Flight Secondary Ion Mass Spectrometry
(g)
Accelerator Mass Spectrometry
(c)
Particle-Induced X-ray Emission
Aerosol Time-of-Flight Mass Spectrometry
(d)
Method 1
Method 2

Sample Number	Year Produced	^{14}C date	Additives (Y/N)
1	1963	1955	N

FIGURE 13.16 Summary of the results of the application of these non-DNA-based techniques to *Bacillus* bacterial spores. With the exception of atomic force microscopy, which was performed on a nonweaponized sample of *Bacillus globigii*, all of the other methods analyzed a single weaponized (powdered) surrogate sample of *Bacillus globigii*. (See color insert.)

observed. While the spores themselves appeared relatively clean (Fig. 13.16b), we observed what we believe to be clumps of silica mixed in with the spores. The EDX spectra confirmed the presence of silica in the sample. EDX analysis provided a good measure of the low Z elements—carbon through calcium. In this sample, carbon, nitrogen, oxygen, sodium, magnesium, aluminum, phosphorus, sulfur, potassium, and calcium were also identified, as shown in the representative spectra in Figure 13.16b. We attempted to quantify the amounts of elements with EDX but found that the nonplanar shapes of our sample created difficulty for the existing system.

To get a quantitative measure of all elements above sodium in the sample, we used nuclear microscopy (PIXE) to determine elemental composition. PIXE was clearly found to be a necessary technique for determining elemental distribution where spatial distribution of elements is required (i.e., measuring the elemental composition of a single spore in a sample). The powder was again dusted onto a sample platform of a nylon membrane, and no coatings of any type were used. An analysis of a single *Bacillus* spore revealed femtogram (i.e.,

10^{-15} gm) quantities of several elements including silica, phosphorus, sulfur, chlorine, potassium, calcium, manganese, iron, and zinc (Fig. 13.16c). It should be noted that these PIXE results corroborate the EDX's finding of silica, phosphorus, sulfur, potassium, and calcium while also determining the presence of additional elements. Interestingly, PIXE results of the bulk sample revealed the presence of titanium (data not shown), a finding for which we have not yet determined a correlation in sample growth or processing.

Bioaerosol mass spectrometry results (see Fig. 13.16d) of the powdered sample yielded identification of the particles as *Bacillus* spores, and we were able to correlate the mass spectra with a method of processing. It is important to realize that the mass spectral signature of an agent particle will reflect both the intrinsic biological agent and the matrix material in which it is embedded or coated. Therefore, the data generated by this instrument contain weaponization signatures as well as organism identification. To analyze the sample by aerosol mass spectrometry, we aerosolized ~0.5 mg of the powder and introduced the particles into the instrument through a converging nozzle. BAMS results (see Fig. 13.16d, left bar) indicate a preponderance of *Bacillus* spores without the presence of chlorine in this powder sample. A second powder (second bar from the left) was shown to contain chlorine in addition to spores. In all, six samples were investigated and grouped by method of processing (e.g., method I, method II). BAMS was able to determine that those made by method I contained spores alone, while those manufactured by method II were found to be rich in chlorine as well (see rightmost bars in Fig. 13.16d). The significance of this is that a BAMS detector can clearly provide process related data that may be useful for attribution. In cases where a covert dissemination takes place, and samples of the bioagent may not be easily recoverable, the BAMS system might provide the *only* data.

Results of surface analysis and depth profiling using time-of-flight secondary ion mass spectrometry are shown in Figure 13.16e. Powder was dusted onto either a beryllium or silicon wafer in order to provide a flat substrate for ion sputtering. The sample was sputtered using a 1 keV cesium ion beam and then imaged with a 25 keV gallium ion beam. As was the case with the BAMS system, the ToF-SIMS spectra also contained weaponization signatures. The difference is that, whereas the BAMS signature is for the entire particle, the ToF-SIMS signatures can be obtained as a function of depth through the sample. Chemical maps generated from the spore surface indicated the presence of sodium, potassium, and chlorine, as expected. Mass spectra of the surface molecular fragments show the presence of silica over and above the background from a silicon substrate, as well as negative ions generated from surface proteins (i.e., CNO^-, CN^-). (The negative ions resulting from proteins are also seen with the bioaerosol mass spectrometer, providing a corroboration of results.) A depth profile through the center of the spore shows a slightly increasing presence of sodium, potassium, and protein fragments (i.e., CNO^-, CN^-) as the center of the spore is reached, with a concomitant increase in calcium. An interesting finding using ToF-SIMS was the increasing amount of fluorine in the spore as the center was approached (Fig. 13.16e). Fluorine may have been naturally present in the water used to grow the initial bacterial sample.[71] These ToF-SIMS results of elemental composition complement those of EDX and PIXE; again, there is overlap (i.e., oxygen, sodium, potassium, phosphorus, calcium, chlorine, silica), as well as unique findings (i.e., fluorine). It should be noted that this ToF-SIMS technique with its ~100-nm spot size allows identification of the spatial location of the element within the spore.

Raman spectroscopy from the sample produced spectra similar to those generated by nonweaponized *Bacillus* spores, again confirming that the Raman signal is due to the presence of calcium-dipicolinate in the spore (Fig. 13.16f). To analyze the powder, it was first dissolved into distilled water before drying a drop of the solution on a CaF_2 puck for analysis. Without this water dissolution step, we found that the powder itself generated too much fluorescence and the particles were too convoluted to identify a single spore for analysis. Results discussed earlier of Raman and surface-enhanced Raman spectroscopy using nonweaponized spores (see the section in this chapter entitled "Raman and Surface-Enhanced Raman Spectroscopy") showed that SERS can distinguish between species of spores by probing the first few nanometers of the spore surface. This technique may prove to be an important tool for forensic analyses because of its high spatial resolution, good chemical selectivity, and high sensitivity.

In order to determine the age of the powder, accelerator mass spectrometry was applied to measure the ^{14}C content of the sample, thereby determining a date based on the amount of that isotope present. As has been pointed out earlier (in the section describing Accelerator Mass Spectrometry), AMS analyses have demonstrated that precise ^{14}C measurements can be determined from relatively small samples. These samples were taken from 1950s archives and showed the incorporation of nuclear testing-derived ^{14}C, yielding dating results consistent with their production dates. For the specific powdered samples investigated here, less than 1 mg of the sample powder was combusted to graphite and analyzed by the 10 MeV accelerator mass spectrometer; the result for one such sample is shown in Figure 13.16g. The date arrived at by ^{14}C analysis was 1955, in reasonable agreement with the actual date produced of 1963. The actual measurement of ^{14}C in the sample is very accurate, but its interpretation is less so. The time resolution of ^{14}C dating is ~1–2 years, and is somewhat less for more recent samples due to the flattening of the "bomb curve" which measures the ^{14}C content of the atmosphere. In this case, the older ^{14}C-derived date may indicate that the medium used to grow the sample was old or that there were other influences due to sample storage and handling later on. There are several factors which can change the ratio of ^{14}C and influence a ^{14}C-derived date. For example, the addition of fossil fuel carbon can lower the amount of ^{14}C in a sample and make the sample appear "younger" than it actually is, while ^{14}C contamination increases the ^{14}C content and artificially "ages" the sample. Fractionation of carbon isotopes by bacteria is potentially another confounding factor, albeit less well understood for ^{14}C than other isotopes.

Clearly there is a need for methods to date the manufacture of BW agents that can be applied to very recent samples, and are independent of some of the possible artifacts that can influence ^{14}C dating.

DISCUSSION AND FUTURE DIRECTIONS

We have demonstrated that there are many promising and unique techniques which can be applied to the analysis of biological agents to provide biological signatures with high resolution and high sensitivity, including both "bulk" or whole-spore analysis and surface analysis. As an example, we have described their application to a powdered weaponized biological surrogate.

Other high-resolution techniques are currently being applied to this investigation. Environmental scanning electron microscopy (ESEM) is also available for taking high-resolution images under hydrated conditions; this has the advantage of avoiding any artifacts associated with the critical point drying process that is required under normal SEM operations. The ESEM is also equipped with EDX and Backscatter capabilities. Fourier transform mass spectrometry using our 9.4 Tesla magnet is being employed to perform studies of the spore surface and the compounds which have adsorbed to it. Similarly, our nano-SIMS (secondary ion mass spectrometer) will enable studies of elements and their isotopic ratios at ~50 nm resolution.

Because we now have the capability to analyze a single spore or biological entity, it is crucial that our samples be grown in a systematic and homogeneous manner. For techniques which analyze a "bulk population," individual differences between cells or spores in that population are averaged out and the resultant data reflect that average population. However, our ability to analyze a single entity allows resolution of differences between individual cells in a population. Our signature libraries will depend upon highly controlled, systematic growth and processing conditions to produce biological entities to be analyzed by these techniques to identify unique signatures. In the process, large quantities of data will be generated, similar to what is already being "mass produced" by DNA sequencing and gene analysis. Therefore, algorithms need to be developed and implemented for each of these techniques to support the large databases that each technique can and will generate.

As analyses by these techniques continue and signature libraries of correlations between those analyses and growth and processing conditions grow, it will be necessary to develop an information system which combines disparate types of data to determine unique signatures. For example, the system would link weapons manufacturing information and associations with foreign or domestic terrorist groups, weaponization signatures as described in this chapter, DNA sequences, component materials analyses, and metabolic signatures. The information system must provide a transparent and flexible query structure for comparison of new unknown samples with existing data, available for any type of data from SEM images to Raman spectra. Algorithms capable of combining several signatures or pieces of data from individual techniques and generating a distinguishing signature for a growth process will need to be developed.

We have demonstrated that multiple high-resolution, high-sensitivity techniques can and must be applied on the single-spore level. This is the path necessary to link the "how, when, and where" of a bioagent to the perpetrator of a biocrime.

REFERENCES

1. Baden M, Roach M. *Dead Reckoning: The New Science of Catching Killers*. New York, NY: Simon & Schuster; 2001.
2. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. AFLP: a new technique for DNA fingerprinting. *Nucl Acids Res* 1995;23:4407–4414.
3. Jackson PJ, Hill KK, Laker MT, Ticknor LO, Keim P. Genetic comparison of *Bacillus anthracis* and its close relatives using amplified fragment length polymorphism and polymerase chain reaction analysis. *J Appl Microbiol* 1999;87:263–269.
4. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson, PJ, Hugh-Jones ME. Multiple-Locus Variable-Number Tandem Repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol* 2000;182:2928–2936.
5. Dang JL, Heroux K, Kearney J, Arasteh A, Gostomski M, Emanuel PA. *Bacillus* spore inactivation methods affect detection assays. *Appl Environ Microbiol* 2001;67:3665–3670.
6. Binning G, Quate CF, Gerber C. Atomic force microscope. *Phys Rev Lett* 1986;56:930–933.
7. Hansma PK, Cleveland JP, Radmacher M, Walters DA, Hillner PE, Bezanilla M, Fritz D, Vie D, Hansma HG, Prater CB, Massie J, Fukunage L, Gurley J, Elings V. Tapping mode atomic force microscopy in liquids. *Appl Phys Lett* 1994;64:1738–1740.
8. Muller DJ, Amrein M, Engel A. Adsorption of biological molecules to a solid support for scanning probe microscopy. *J Struct Biol* 1997;119:172–188.
9. Muller DJ, Schoenenberger CA, Schabert F, Engel A. Structural changes in native membrane proteins monitored at subnanometer resolution with the atomic force microscope: A review. *J Struct Biol* 1997;119:149–157.
10. Lyubchenko YL, Schlyakhtenko LS. Visualization of supercoiled DNA with atomic force microscopy in situ. *Proc Natl Acad Sci USA* 1997;94:496–501.
11. Han W, Lindsay SM, Dlakic M, Harrington RE. Kinked DNA. *Nature* 1997;386:563.

12. Bustamante C, Rivetti C. Visualizing protein-nucleic acids interactions on a large scale with the scanning force microscope. *Annu Rev Biophys Biomol Struct* 1996;25:395–429.
13. Kuznetsov YG, Malkin AJ, Lucas RW, Plomp M, McPherson A. Imaging of viruses by atomic force microscopy. *J Gen Virology* 2001;82:2025–2034.
14. Plomp M, Rice MK, Wagner EK, McPherson A, Malkin AJ. Rapid visualization at high resolution of pathogens by atomic force microscopy: Structural studies of Herpes Simplex Virus-1. *Am J Pathology* 2002;160:1959–1966.
15. Malkin AJ, McPherson A, Gershon PD. Structure of intracellular mature vaccinia virus visualized by in situ AFM. *J Virology* 2003;77:6332–6340.
16. Canady MA, Larson SB, Day J, McPherson A. Crystal structure of turnip yellow mosaic virus. *Nat Struct Biol* 1996;3:771–781.
17. Malkin AJ, Kuznetsov YG, Lucas RW, McPherson A. Surface processes in the crystallization of turnip yellow mosaic virus visualized by atomic force microscopy. *J Struct Biol* 1999;127: 35–43.
18. Baker TS, Olson NH, Fuller SD. Adding the third dimension to virus life cycles: Three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbol Mol Biol Reviews* 1999;63:862–922.
19. Dai H, Hafner JH, Rinzler AG, Colbert DT, Smalley RE. Nanotubes as nanoprobes in scanning probe microscopy. *Nature* 1996;384:147–151.
20. Hafner JH, Cheung CL, Lieber CM. Growth of nanotubes for force microscopy tips. *Nature* 1999;398:761–762.
21. Hafner JH, Cheung CL, Woolley AT, Lieber CM. Structural and functional imaging with carbon nanotube AFM probes. *Progr Biophys Mol Biol* 2001;77:73–110.
22. Hafner JH, Cheung CL, Oosterkamp TH, Lieber CM. High-yield assembly of individual single-walled carbon nanotube tips for scanning probe microscopies. *J Phys Chem* 2001;B105: 743–746.
23. Raman CV, Krishnan KS. A new type of secondary radiation. *Nature* 1928;121:501.
24. Raman CV. A change of wavelength in light scattering. *Nature* 1928;121:619.
25. Lewis IR, Edwards HGM (eds). *Handbook of Raman Spectroscopy. Practical Spectroscopy, Vol. 28*. New York: Marcel Dekker; 2001.
26. Twardowski J, Anzenbacher P, Raman CV. IR spectroscopy in biology and biochemistry. *In* Masson M, Tyson JF, Stockwell P (eds). *Ellis Horwood Series in Analytical Chemistry*. London: Ellis Horwood; 1994.
27. Spiro TJ. *Biological Applications of Raman Spectroscopy*. New York, NY: John Wiley & Sons; 1989.
28. Naumann D. FT-infrared and FT-Raman spectroscopy in biomedical research. *Appl Spectrosc Rev* 2001;36:239–298.
29. Choo-Smith LP, et al. Medical applications of Raman spectroscopy: From proof of principle to clinical implementation. *Biopolymers* 2002;67:1–9.
30. Nelson WH, Manoharan R, Sperry JF. UV resonance Raman studies of bacteria. *Appl Spectrosc Rev* 1992;27:67–124.
31. Shibata H, et al. Laser Raman spectroscopy of lyophilized bacterial spores. *Microbiol Immunol* 1986;30:307–313.
32. Ghiamati E, et al. UV resonance Raman spectra of bacillus spores. *Appl Spectrosc* 1992;46:357–364.
33. Esposito AP, Talley CE, Huser T, et al. Analysis of single bacterial spores by micro-Raman spectroscopy. *Appl Spectrosc* 2003;57:868–871.
34. Kneipp K, Kneipp H, Itzkan I, et al. Ultrasensitive chemical analysis by Raman spectroscopy. *Chem Rev* 1999;99:2957–2975.
35. Kneipp K, Wang Y, Kneipp H, et al. Single molecule detection using surface-enhanced Raman scattering (SERS). *Phys Rev Lett* 1997;78:1667–1670.

36. Nie S, Emory SR. Probing single molecules and single nanoparticles by surface-enhanced Raman scattering. *Science* 1997;275:1102–1106.
37. Suess DT, Prather KA. Mass spectrometry of aerosols. *Chemical Reviews* 1999;99:3007–3035.
38. Gard EE, Kleeman MJ, Gross DS, Hughes LS, Allen JO, Morrical BD, Fergenson DP, Dienes T, Galli ME, Johnson RJ, Cass GR, Prather KA. Direct observation of heterogeneous chemistry in the atmosphere. *Science* 1998;279:1184–1187.
39. Gard EE, Mayer JE, Morrical BD, Dienes T, Fergenson DP, Prather KA. Real-time analysis of individual atmospheric aerosol particles—Design and performance of a portable ATOFMS. *Anal Chem* 1997;69:4083–4091.
40. Song XH, Hopke PK, Fergenson DP, Prather KA. Classification of single particles analyzed by ATOFMS using an artificial neural network, ART-2A. *Anal Chem* 1999;71:860–865.
41. Phares DJ, Rhoads KP, Wexler AS, Kane DB, Johnston MV. Application of the ART-2a algorithm to laser ablation aerosol mass spectrometry of particle standards. *Anal Chem* 2001;73:2338–2344.
42. Trimborn A, Hinz KP, Spengler B. Online analysis of atmospheric particles with a transportable laser mass spectrometer. *Aerosol Sci Technol* 2000;1-2:191–201.
43. Hinz KP, Greweling M, Drews F, Spengler B. Data processing in on-line laser mass spectrometry of inorganic, organic, or biological airborne particles. *J Am Soc for Mass Spect* 1999;10:648–660.
44. Benninghoven A, Ruedenauer FG, Werner HW. *Secondary Ion Mass Spectrometry Basic Concepts, Instrumental Aspects, Applications and Trends*. New York, NY: John Wiley & Sons; 1987.
45. Vickerman JC, Brown A, Reed NM. *Secondary Ion Mass Spectrometry,* Vol. 17. Oxford, UK: Clarendon Press; 1989.
46. Briggs D, Brown A, Vickerman JC. *Handbook of Static Secondary Ion Mass Spectrometry.* New York, NY: John Wiley & Sons; 1989.
47. Schueler BW. Microscope imaging by time-of-flight secondary ion mass spectrometry. *Microsc Microanal Microstruct* 1992;3:119–140.
48. Delcorte A, Bertrand P, Wischerhoff E, Laschewsky A. Adsorption of polyelectrolyte on polymer surfaces. *Langmuir* 1997;13:5125–5236.
49. Arlinghaus HF, Hueppener C, Drexler J. TOF-SIMS characterization of DNA and PNA biosensor chips. *In* Benninghoven A, Bertrand P, Migeon HN, Werner HW (eds). *Secondary Ion Mass Spectrometry XII.* Amsterdam, The Netherlands: Elsevier Science; 2000:951–954.
50. Walker AV, Winograd N. Prospects for imaging with TOF-SIMS using gold liquid metal ion sources. *Applied Surf Sci* 2002:9233–9234.
51. Wu KJ, Odom RW. Matrix-enhanced secondary ion mass spectrometry: A method for molecular analysis of solid surfaces. *Anal Chem* 1996;68:873–882.
52. Wagner MS, Castner DG. Characterization of adsorbed protein films by time-of-flight secondary ion mass spectrometry with principal component analysis. *Langmuir* 2001;17: 4649–4660.
53. Wagner MS, Tyler BJ, Castner DG. Interpretation of static time-of-flight secondary ion mass spectra of adsorbed protein films by multivariate pattern recognition. *Anal Chem* 2002;74:1824–1835.
54. Todd PJ, McMahon JM, Short RT, McCandlish CA. Organic SIMS of biological tissues. *Anal Chem* 1997;69:529A–535A.
55. Pacholski ML, Winograd N. Imaging with mass spectrometry. *Chem Rev* 1999;99:2977–3005.
56. Roddy TP, Cannon DM, Meserole CA, Winograd N, Ewing AG. Imaging of freeze-fractured cells with in situ fluorescence and time-of-flight secondary ion mass spectrometry. *Anal Chem* 74:4011–4019.
57. Cannon DM, Pacholski ML, Winograd N, Ewing AG. Molecule specific imaging of freeze-fracture frozen-hydrate model membrane systems using mass spectrometry. *J Am Chem Soc* 2000;122:603–610.

58. Cahill TA, Cliff SS, Jimenez-Cruz M, Shackelford JF, Dunlap M, Meier M, Kelly PB, Riddle S, Selco J, Bench G, Grant PG, Ueda D, Perry KD, Leifer R. Analysis of Aerosols from the World Trade Center Collapse Site, New York, October 2 to October 30, 2001, *Aerosol Sci Technol* 2004;38:165–183.
59. Bench G, Carlson TM, Grant PG, Wollett JS Jr, Martinelli RE, Lewis JL, Divine KK. Olfactory bulb uptake and determination of biotransfer factors in the California ground squirrel (*Spermophilus beecheyi*) exposed to manganese and cadmium in environmental habitats. *Environ Sci Technol* 2001;35:270–277.
60. Vogel JS, Grant PG, Buchholz BA, Dingley K, Turteltaub KW. Attomole quantification of protein separations with accelerator mass spectrometry. *Electrophoresis* 2001;22:2037–2045.
61. Hu Z, Bonifas JM, Beech J, Bench G, Shigihara T, Ogawa H, Ikeda S, Mauro T, Epstein HE. Mutations in ATP2C1, encoding a calcium pump, cause Hailey-Hailey disease. *Nature Genetics* 2000;24:61–65.
62. Bench G, Corzett MH, De Yebra L, Oliva R, Balhorn R. Protein and DNA contents in sperm from an infertile human male possessing protamine defects that vary over time. *Mol Reprod Devel* 1998;50:345–353.
63. Roberts ML, Grant PG, Bench GS, Brown TA, Frantz BR, Morse DH, Antolak AJ. The stand-alone microprobe at Livermore. *Nucl Instr & Meth* 1999;B158:24–30.
64. Levin I, Kromer B. Twenty years of atmospheric $^{14}CO_2$ observations at Schauinsland Station, Germany. *Radiocarbon* 1997;39:205.
65. Stuiver M, Polach HA. Discussion: Reporting of ^{14}C data. *Radiocarbon* 1977;19:355.
66. Vogel JS, Nelson DE, Southon JR. ^{14}C background levels in an accelerator mass spectrometry system. *Radiocarbon* 1987;29:323.
67. Vogel JS. Accelerator mass spectrometry as a bioanalytical tool. *In* Heys JR, Melillo DG (eds). *Synthesis and Applications of Isotopically Labelled Compounds*. New York, NY: John Wiley & Sons; 1997:11–19.
68. Plomp M, Leighton TJ, Wheeler KE, Malkin AJ. The high-resolution architecture and structural dynamics of *Bacillus* spore. *Biophys J* 2005;88:603–608.
69. Dufrene YF, Boonaert CJP, Gerin PA, Asther M, Rouxhet PG. Direct probing of the surface ultrastructure and molecular interactions of dormant and germinating spores of *Phanerochaete chrysosporium*. *J Bacteriol* 1999;181:5350–5354.
70. van der Aa BC, Asther M, Dufrene YF. Surface properties of *Aspergillus oryzae* spores investigated by atomic force microscopy. Colloids surfaces B. *Biointerfaces* 2002;24:277–284.
71. Wu KJ. Manuscript in preparation.

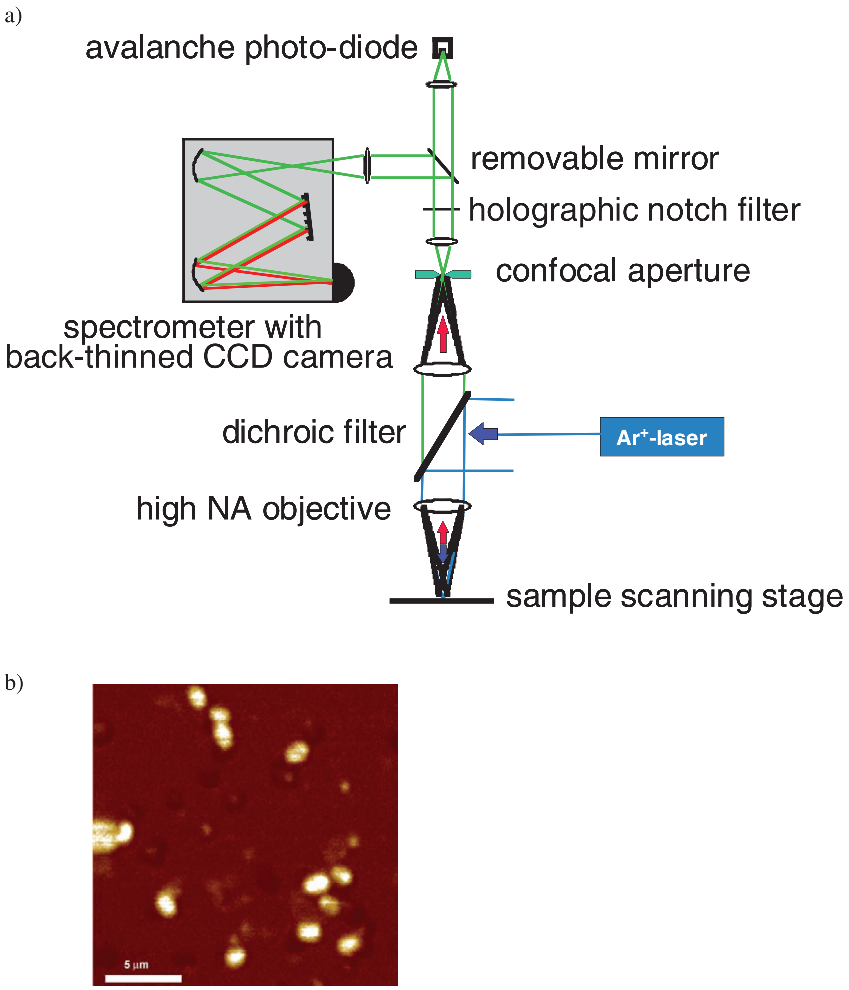

a)
avalanche photo-diode
removable mirror
holographic notch filter
confocal aperture
spectrometer with
back-thinned CCD camera
dichroic filter
Ar+-laser
high NA objective
sample scanning stage
b)
5 µm

FIGURE 13.8 Schematic illustration of confocal microscope used for the micro-Raman experiments. (A) The 488-nm line of an Ar^+ laser is focused onto the sample with a high numerical aperture (NA) objective. The sample is mounted on an xy-piezo stage and raster-scanned over the focused laser beam. Raman scatter is collected with the same high-NA objective and focused onto a confocal aperture. The Rayleigh scatter is removed with a holographic notch filter, and the remaining Raman scatter is focused onto an avalanche photo-diode, the signal which is used to build up the image. Once a feature of interest is located, the laser beam is centered on the object and the Raman scatter is sent into a spectrometer with a back-thinned CCD camera to acquire the Raman spectra. (B) Confocal image of individual bacterial spores dried onto a calcium fluoride substrate. The bright spots correspond to the intrinsic autofluorescence and Raman scatter from the spores. Once the autofluorescence is photobleached, the Raman spectrum from the individual spore is collected.

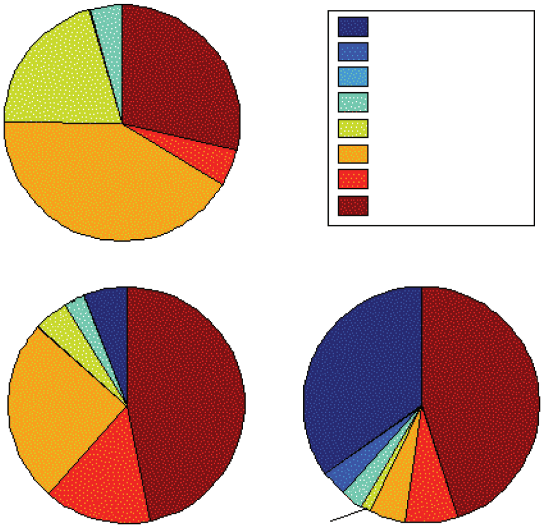


FIGURE 13.11 BAMS at work: Three samples of particles analyzed by our real-time spectrum identification software. (a) Only the background materials have been added. (b) In addition to the background materials, *Bacillus* spores have been added. (c) In addition to the background materials and *Bacillus* spores, *Clostridium* spores have been added.

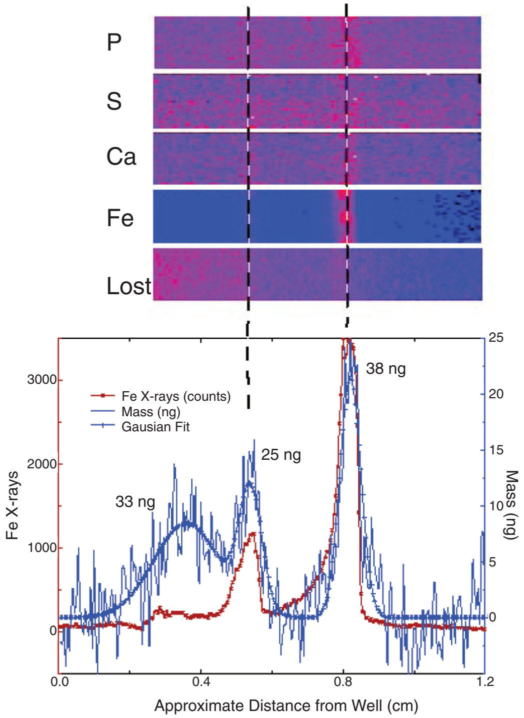


FIGURE 13.14 Location of ferritin in an electrophoretic gel. The figure is oriented such that the gel's well is on the left. The top portion of the figure shows the location of several elements in the gel as identified by PIXE analysis. The bottom graph indicates the co-location of proteins and iron in the gel. Iron (red line) was measured by PIXE simultaneously with STIM measurements of protein mass (blue line). Ferritin, an iron-binding protein, is shown at ~0.8 cm from the well, the overlap of the protein and iron curves.

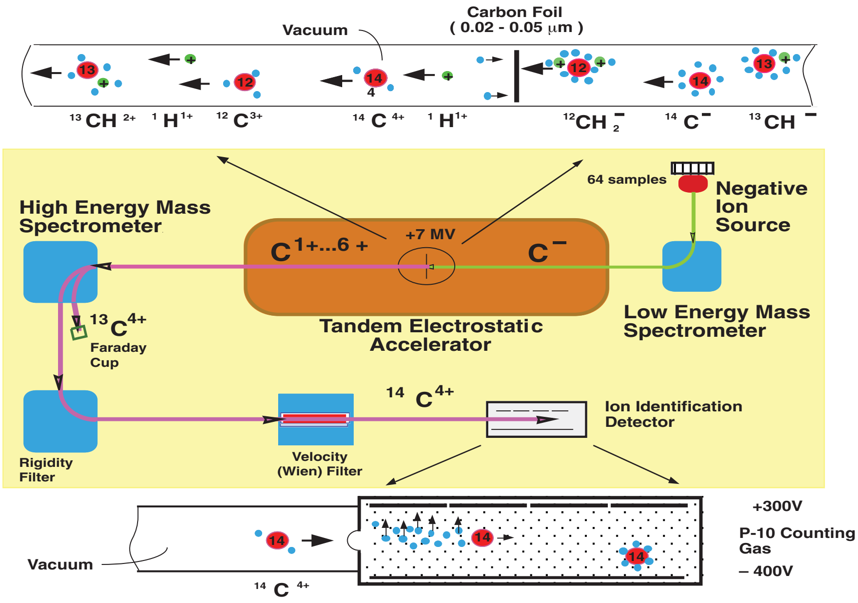

Vacuum
Carbon Foil
(0.02 - 0.05 µm)
13
12
14
4
^{13}CH $^{2+}$
^{1}H^{1+}
^{12}C^{3+}
^{14}C $^{4+}$
^{1}H^{1+}
^{12}CH$_2^-$
^{14}C$^-$
^{13}CH $^-$
64 samples
Negative Ion Source
High Energy Mass Spectrometer
C$^{1+...6+}$
+7 MV
C$^-$
Tandem Electrostatic Accelerator
Low Energy Mass Spectrometer
^{13}C^{4+}
Faraday Cup
^{14} C^{4+}
Ion Identification Detector
Rigidity Filter
Velocity (Wien) Filter
Vacuum
^{14} C $^{4+}$
+300V
P-10 Counting Gas
– 400V

FIGURE 13.15 Schematic of the 10 MeV accelerator mass spectrometry system.

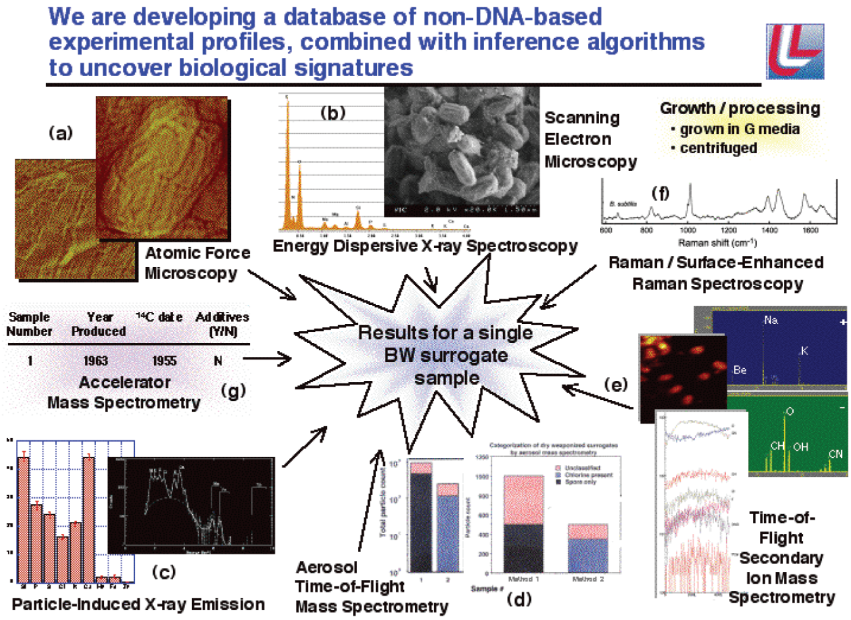

We are developing a database of non-DNA-based experimental profiles, combined with inference algorithms to uncover biological signatures
(a)
Atomic Force Microscopy
(b)
Energy Dispersive X-ray Spectroscopy
Scanning Electron Microscopy
Growth / processing
• grown in G media
• centrifuged
(f)
Raman shift (cm⁻¹)
Raman / Surface-Enhanced Raman Spectroscopy
Results for a single BW surrogate sample
Sample Number | Year Produced | ¹⁴C date | Additives (Y/N)
1 | 1963 | 1955 | N
Accelerator Mass Spectrometry
(g)
(e)
Na
K
Be
O
CH
OH
CN
(c)
Particle-Induced X-ray Emission
Aerosol Time-of-Flight Mass Spectrometry
Total particle count
Categorization of dry weaponized surrogates by aerosol mass spectrometry
Unclassified
Chlorine present
Spore only
Particle count
Method 1
Method 2
Sample #
(d)
Time-of-Flight Secondary Ion Mass Spectrometry

FIGURE 13.16 Summary of the results of the application of these non-DNA-based techniques to *Bacillus* bacterial spores. With the exception of atomic force microscopy, which was performed on a nonweaponized sample of *Bacillus globigii*, all of the other methods analyzed a single weaponized (powdered) surrogate sample of *Bacillus globigii*.

CHAPTER 14

Microbial Forensics Host Factors

STEVEN E. SCHUTZER

Department of Medicine, University of Medicine and Dentistry of New Jersey—New Jersey Medical School, Newark, New Jersey

Consider the diagnosis of a disease caused by a biothreat. Multiple questions arise immediately. One question is whether it is a natural event or an intentional attack. Knowledge of the host response can potentially answer some questions depending upon the pathogen involved. Some of these questions include who else was exposed, who else may have been vaccinated for protection, how long ago was the person infected, and were other persons who are close to the individual also infected before or after the index case? Aside from public and personal health concerns, these questions seek to categorize individuals as possible victims versus perpetrators. It is the aim of this chapter to present the elements of the host response in a simplified fashion that in the right context has high potential to bring answers to these questions.

Microbial forensics has a primary aim to identify the biological agent, its source, and the individuals responsible for a biothreat event.[1] Analytic approaches differ when the suspected biothreat agent is encountered in a container or the environment, as opposed to a human or animal. Trace element, pollen, growth media, latent fingerprint, and microbial and nonmicrobial nucleic acid analysis are all applicable to the container and environmental sample.[2] However, once the microbe or its toxin is in the living host, it is no longer possible to analyze all of the preceding items except the microbial nucleic acid. Nevertheless, the host response to the biological agent is available for analysis. This is akin to other forensic studies where physical traces of bite marks, scratches, wound trajectories, and sizes of wounds are often surrogate evidence of the teeth, fingernails, and bullets.[3] The forensic pathologist is already familiar with these. Those involved with epidemiologic and diagnostic issues will be more familiar with the host response. In the context of microbial forensics it is important to integrate all of these with intelligence information so that an authenticated piece of a puzzle may be included in the analytical and attribution picture.

Microbial Forensics
Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

The host response to a foreign substance is often a well orchestrated series of events designed to protect the individual from harm. Modern techniques help us elucidate the pathways and components of the host response. The immune system and its components are a mainstay of our protection against infections and malignancies.[4,5] Inflammation is often an unpleasant side effect as the immune system contains and eradicates a microbe or foreign tissue. Specific arms of the immune system can be used as markers in favor or against the presence of an infection. The humoral or antibody response to an invading microbe is an example. Some of the antibodies that are produced have a protective effect with other parts of the immune-inflammatory system and are responsible for eradicating the infection. Other antibodies may not be as effective in this role. However, in their ability to recognize unique and specific structures of a microbe, they serve as beacons that a microbe was recently present or was present in the distant past. Substances such as antibiotics which can rapidly kill a microbe may modify the immune response by removing the infectious driving force for a full-scale response. In clinical medicine and veterinary medicine, measurement of the immune response helps the diagnostician decide what infection was present and how recently. In these situations the intention is to provide treatment. For other pieces of the puzzle, the forensic scientist may exploit parts of the immune response to discover who is likely a victim of an attack and who might be responsible. This chapter will discuss the basics of the host immune response that can have utility in the microbial forensic sense. Examples will provide a sense of what information is achievable and what is not likely to provide clues with a high degree of certainty.

GENERAL CONCEPTS

In response to an encounter with a new microbe, the immune system first starts to activate the antibody system. Usually a cell known as a macrophage engulfs some of the microbes. It then presents part of the microbe to a helper T cell (a lymphocyte) which then directs other lymphocytes known as B cells to produce antibodies to that particular microbe and even more specifically to that part which was presented. It usually takes at least 4 days before any microbe-specific antibody can be found.[6]

Antibodies are a specific form of the proteins known as immunoglobulins (Igs). IgM, IgG, and IgA are the principal classes of immunoglobulins with relevance to this chapter and will be discussed in more detail. Those individuals unfortunate to have allergies have problems due to IgE against allergens (such as ragweed, peanut, or cat dander). In this case the IgE molecules sit on the surface of cells that can release histamine when the offending allergen bridges

TABLE 14.1 Immunoglobulin classes and properties

Immunoglobulin Class	IgM	IgG	IgA	IgE	IgD
Size (kD)	900	150	160	190	180
Serum half-life days	5	21–23	5–6	1–5	2–8
Placental transfer	No	Yes	No	No	No
Complement fixation	++	+	−	−	−
Percentage of serum immunoglobulin	80	13	6	0.002	0.2

two such molecules. In an infection, immunoglobulins usually appear in the order of IgM, IgG, and IgA. B cells first begin to produce IgM, and then some B cells undergo an irreversible switch to those that produce IgG. Later some of this population of cells undergo a switch to become IgA-secreting type B cells. Immunoglobulins persist for varying times; for example, the half-life of particular IgM antibodies is 5 days, while that of IgG can be as long as 21–23 days (Table 14.1).[4,6]

Similar to a live microbe, vaccines can also provoke an antibody response. The vaccine can be composed of a live or attenuated microbe, a whole non-proliferating microbe, or an antigenic part of the microbe. Regardless, the intent of the vaccine is to produce protection, often by protective antibodies. Although the half-life of an individual IgG molecule is less than a month, a population of antibodies in the IgG isotype form may persist for life. Memory B cells can sustain these antibodies and retain the ability to quickly generate the appropriate antibodies when challenged. When the immune system encounters another infection or is subjected to a revaccination (booster), the result is an accelerated production of the particular antibody and increase in the levels that circulate in the blood. Figure 14.1 illustrates this.

Perhaps the most discernible pattern of antibody response which has forensic value is the appearance of IgM first, followed by a B-cell switch to the longer lasting IgG. During the early phase of exposure, IgM can be seen first. As time goes on, IgG is seen and predominates, and IgM is no longer found. This is illustrated in Figure 14.2.

The antibody response to a particular agent may be directed to different antigens at different times, that is, early or later after the initial exposure. That response may involve IgM at the early stage and IgG later. Late in the disease or during recovery, only IgG to particular antigens may be seen. A classic example is the human antibody response to Epstein Barr virus (EBV).[7] EBV is the virus known to cause mononucleosis. During acute early disease, it is common to find high levels of antibody of the IgM isotype to the viral early

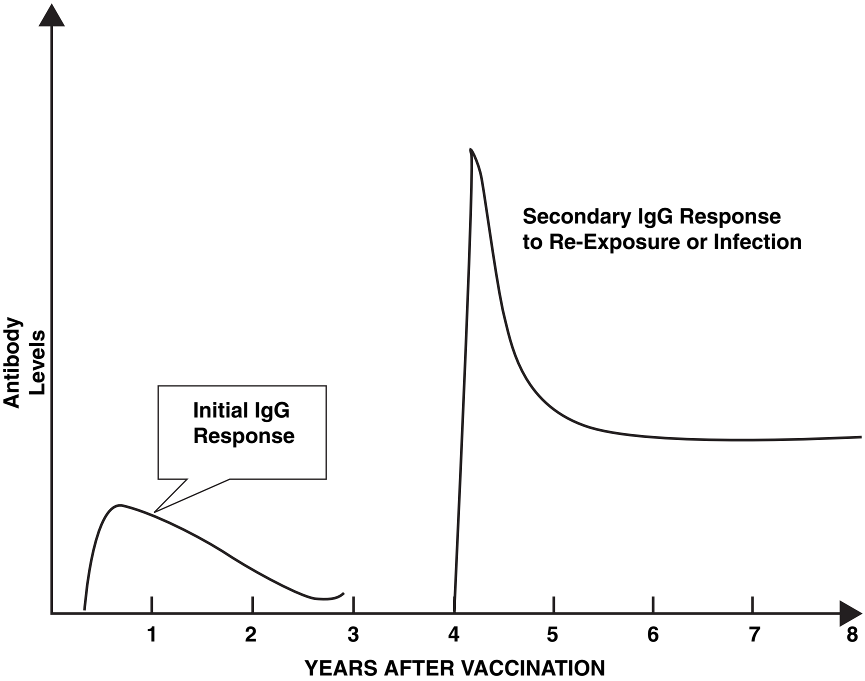


FIGURE 14.1 Illustration of the IgG antibody response to a vaccine antigen after the first immunization and subsequent exposure by natural exposure to the infectious agent or by another vaccination.

antigen (EA) and viral capsid antigen (VCA). It is rare to find IgG antibody to the VCA or Epstein Barr Nuclear Acid (EBNA) in anything but low titers. As the patient recovers from their first infection with EBV, it is rare to find anything but low levels of IgM to EA or VCA, but IgG to VCA in higher or increasing levels is common. Antibodies to EBNA are often very low during this stage. Then after clinical recovery, that is, several months later, IgM to EA and VCA stay at low levels whereas IgG to VCA and EBNA remain at high levels, often for years. Table 14.3 illustrates this pattern by stage of the immune response to EBV and its particular antigens. Figure 14.3 is a graphic display of this. For the clinician or epidemiologist this provides a framework to determine where in the infectious process a patient may be. Tables 14.2 and 14.3 and Figures 14.2 and 14.3 illustrate how responses to a biothreat agent or its toxin may be used to give some chronological indication of exposure. Combining the antibody response with detection of particular antigens can provide further definition as to the stage of infection or exposure.

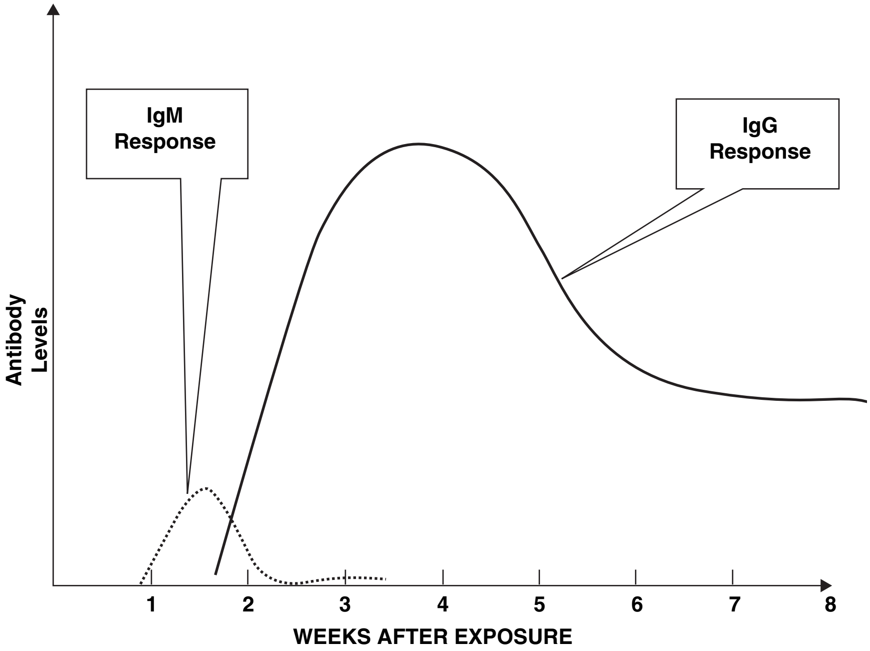


FIGURE 14.2 Illustration of the temporal relation of the IgM and IgG responses to an infection with IgM as the first and often transient response and IgG as the more sustained response.

TABLE 14.2 Antibody tests for Epstein-Barr virus

Stage	Titers
Acute primary infection	
IgM EA and VCA	High
IgG VCA and EBNA	Low
Recovering from primary infection	
IgM EA or VCA	Lower
IgG VCA	Rising
EBNA	Low
After several months	
IgM EA and VCA	Low or normal
IgG VCA and EBNA	Persist at high for several years

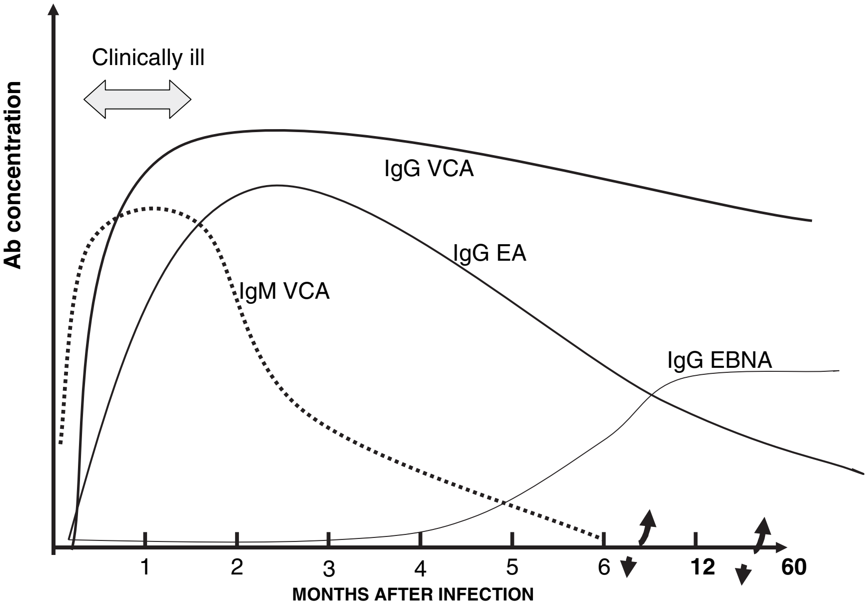


FIGURE 14.3 Schematic response of IgM and IgG to different antigens of EBV over an extended period of time.

TABLE 14.3 Antibody response at different time points to EBV antigens

Disease status	Heterophile Ab	VCA-IgM	VCA-IgG	EBNA	EA(D)
Healthy-unexposed	Negative	Negative	Negative	Negative	Negative
Very early infection	Possible	Possible	Possible	Negative	Negative
Active infection	Positive	Positive	Positive	Negative	Possible
Recent infection	Positive	Positive	Positive	Positive	Possible
Past infection	Negative	Negative	Positive	Positive	Possible

A controlled experiment or normal clinical event illustrates what happens when the immune system sees the infectious agent or its vaccine representation again. The controlled experiment may be in a laboratory animal or a patient receiving a booster vaccine. The uncontrolled but normal clinical event occurs when the patient is exposed again to the infectious agent for whatever reason. Consider a generic antigen exposure. The first time the immune system sees Antigen X (AgX) it responds as shown in Figures 14.1 and 14.2. At first any antibody to AgX is barely discernible; then the levels rise and later fall to a plateau. If a mixed infectious exposure were to occur with AgX and a new AgY from another microbe, the immune system would quickly mount a brisk and high level of Ab to AgX, while the course of Ab to AgY would be slow and delayed, just as was the first exposure to AgX. For AgX this is a phenomenon termed immunological memory or an amnestic response. This can be useful when the symptoms and signs of exposure to either X or Y are similar. This is the case with the early flu-like symptoms of pulmonary anthrax and the influenza virus itself.

Another example common to all of us is repetitive exposure to different strains of flu viruses.[4] As illustrated in Table 14.4, a person infected for the first time with one strain of the influenza virus makes a response to most of its antigens (as a theoretical example, Ag 1, 2, 3, 4, 5, 6). Three years later, the same individual exposed to a partially similar influenza virus responds preferentially to those antigens that were also present on the original influenza virus. The person also makes a smaller initial antibody response to new antigens, that is, those not shared with the first virus. The initial response is minimal in comparison. Ten or 20 years later, during a new flu season and exposure to a third strain of influenza, the most brisk responses would be to antigens previously recognized by the immune system. This is the scientific basis for giving the flu vaccine, which contains a variety of possible antigens common to multiple strains of the flu virus so that a rapid and protective antibody response will occur.

TABLE 14.4 Response to theoretical antigens from different flu viruses at the time of exposure (weak vs. strong)

Infecting strain and antigen composition	Antibody response: Weak	Antibody response: Strong
Strain A Year 1 (Antigens 1, 2, 3, 4, 5, 6)	1, 2, 3, 4, 5, 6	—
Strain B Year 5 (Antigens 1, 3, 5, 7, 8, 9)	7, 8, 9	1, 3, 5
Strain C Year 15 (Antigens 1, 3, 8, 10, 12, 13)	10, 12, 13	1, 3, 8

UTILITY OF SEROLOGIC ANALYSIS OF PEOPLE EXPOSED TO ANTHRAX: STRENGTHS AND LIMITATIONS

Our knowledge of the humoral response to infection with biothreat microbes is limited compared to our knowledge of the kinetics and time response to common human infections. Nevertheless, in the appropriate context and with sufficient background information, detection of antibodies to a particular microbe and its antigens can have important value for microbial forensics. This information may have critical probative value or it can guide investigative leads. The absence of a specific antibody response may also have equal value in a particular investigation. Certainly its importance is increased in the context of knowledge of what organism may be involved, when the exposure was likely to have occurred, the route of exposure, what symptoms and signs are manifesting in the host, and other hard data points such as detection of antigens themselves, and detection of microbial DNA or RNA.[8] Other information such as how many hosts (people or animals) have had this infection in the geographic region, what is the normal infection rate, and background incidence of antibody titer due to the organism in question or a related organism, is also important.

The 2001 anthrax letter attacks raised multiple questions for every person infected, possibly exposed, vaccinated, or treated. Some of these questions included how these persons were infected, if at all; that is, was it by the skin, which could produce cutaneous anthrax; by inhalation of spores, which would produce pulmonary and systemic anthrax infection; did they ingest any spores, which would produce an initial gastrointestinal infection; or, were they among the "worried well?"

Consider the situation where a close associate comes down with symptoms compatible with anthrax infection after receiving a powder-containing letter. Until this is disproved as anthrax, great worry will ensue. We learned that the limited textbook medicine did not apply. Yet there is useful information to be used in the present while designing future studies.

In several cases of documented exposure, there was not enough time for the patient to develop antibody to a specific anthrax antigen, at least as probed for IgG. Serial serum samples obtained on November 16, 17, 18, and 19 of 2001, well after the potential work place exposure during the first week in October 2001, were tested for IgG antibody to the protective antigen (PA) component of the anthrax toxins by enzyme-linked immunosorbent assay (ELISA); all samples were nonreactive. Serial tests for serum IgG antibody to the PA toxin of anthrax were performed on 436 workplace-exposed persons. All but one test was negative. Most specimens were collected on October 10 and 17.[9]

It is instructive to look at the positive antibody case in the context and duration of that individual's symptoms when he developed a positive test. None of the symptoms detailed below were individually unique to raise suspicion of a particular diagnosis. The 73-year-old man (case 2) developed fatigue on September 24. He was the newspaper mailroom clerk who delivered mail to the first patient (case 1). On September 28, he developed a nonproductive cough, intermittent fever, runny nose, and conjunctivitis. These symptoms worsened through October 1 when he was hospitalized. In addition he had shortness of breath with exertion, sweats, mild abdominal pain and vomiting, and episodes of confusion. Temperature was elevated to 38.5°C (101.3°F), heart rate was rapid at 109/min, respiratory rate was slightly fast at 20/min, and blood pressure was 108/61 mm Hg. He had bilateral conjunctival injection and bilateral pulmonary rhonchi. At that time his neurologic exam was normal. No skin lesions were observed. The only laboratory abnormalities were low albumin, elevated liver transaminases, borderline low serum sodium, increased creatinine, and low oxygen content in the blood. Blood cultures were negative on hospital day 2, after antibiotics had been started. The chest X-ray showed a left-sided pneumonia and a small left pleural effusion but no "classical" mediastinal widening. The patient was initially given intravenous azithromycin; cefotaxime and ciprofloxacin were subsequently added. A nasal swab obtained on October 5 grew *Bacillus anthracis* on culture. Computed tomography (CT) of the chest showed bilateral effusions and multilobar pulmonary consolidation but still no significant mediastinal lymphadenopathy. Pleural fluid aspiration was positive for *B. anthracis* DNA by PCR. Bacterial cultures of bronchial washings and pleural fluid were negative. A transbronchial biopsy showed *B. anthracis* capsule and cell-wall antigens by immunohistochemical staining. During hospitalization, the white blood count rose to 26,800/mm,[3] and fluid from a second thoracentesis was positive for *B. anthracis* DNA by PCR. Both pleural fluid cells and pleural biopsy tissue showed *B. anthracis* capsule and cell-wall antigens by immunohistochemical staining. Serial serum samples demonstrated a greater than fourfold rise in serum IgG antibody to the PA component of the anthrax toxins by an ELISA assay. The patient was able to leave the hospital on October 23 on oral ciprofloxacin. Table 14.5 illustrates both the clinical and bioforensic approach and context in which to analyze such a patient. These are likely to be common to most situations where a biocrime is suspected to have affected an individual. The first set of questions revolves around whether the person is sick: does the patient have any indications of not being well and any laboratory evidence suggestive of any infection? The second set of questions addresses whether there is any specific and objective laboratory evidence of a particular infection.

This case points out that direct cultures may be negative at different times from different fluids and tissues. This may be influenced by the early administration of antibiotics. However the remnants of the infection, even dead

TABLE 14.5 Nonspecific and specific indications of a case of anthrax

Clinical evidence of an infection	Nonspecific laboratory evidence of an infection	Specific clinical evidence of an infection with *B. anthracis*
Known exposure by proximity to area and infected person	Chest X-ray and CT scan showing pneumonia and pleural fluid	Culture from nasal swab grew live *B. anthracis*
Cough, fever, shortness of breath	Elevated white blood cell count	Positive PCR for *B. anthracis* in pleural fluid on two occasions despite negative cultures
Sweats, abdominal pain, confusion		Positive immunochemical staining for *B. anthracis* capsule and cell wall antigens of transbronchial biopsy, pleural fluid cells, and pleural biopsy despite negative cultures
Abnormal breath sounds		Serum IgG to PA toxin component
Fast heart rate		Serum IgG titer to PA toxin increased within a short time period

organisms, can be found by probing for antigens and DNA. This patient manifested a classic principle of infectious disease, a rising antibody titer over time. In this case it was IgG to a particular antigen of the anthrax toxin. This antibody response may have been detected earlier if IgM to this toxin or other antigens of anthrax had been sought. The case also points out the utility of integrating the presence of antibody with that of other indications of an anthrax infection. These take their greatest significance during clinical symptoms and signs of infection in a possibly exposed individual.

Early administration of antibiotics can prevent positive cultures from the organism in question. Of the first 10 pulmonary anthrax cases associated with the 2001 anthrax letter attacks, three patients had no culture growth of *B. anthracis* from any clinical samples, but culture was attempted after initiation of antibiotic therapy. The diagnosis was made on the basis of history of exposure in conjunction with compatible symptoms and signs of disease, and objective laboratory findings. *B. anthracis* was identified in pleural fluid, pleural biopsy, or transbronchial biopsy specimens by reactivity with *B. anthracis*-specific cell wall and capsular antibodies, or DNA by PCR on pleural fluid or blood.[10]

It is very important to understand the limitations of the assay used. An IgG-based ELISA against the anthrax toxin's protective antigen (PA) component illustrates the importance of understanding the limitations of an assay. The ELISA assay was developed at the U.S. Army Medical Research Institute of Infectious Disease (USAMRIID) and put into operation after optimization at the CDC[11] for functional sensitivity and specificity for detecting antibody response to *B. anthracis* infection. Its major limitation is its restriction to one antigen and to IgG. Therefore a negative result at the time of early exposure may in effect yield a false-negative result. This identifies a gap in our knowledge and application that can be filled by development of an IgM assay, and perhaps one that is enhanced by probing for other *B. anthracis* antigens or epitopes yet to be discovered.

The assay may be very useful in its present form to screen asymptomatic people with possible exposure. The study by Dewan et al. gives a sense of this, and provides a contemporary background database on a group of individuals who may have been exposed to *B. anthracis*.[12] They evaluated postal workers. Beginning on October 29, 2001, 1,657 employees and others who had been to the Washington, D.C. postal facility went to the D.C. General Hospital for antibiotics additional to those begun on October 21. Serum samples were obtained from 202 individuals, and all were negative for specific IgG antibody to the PA IgG, including the three participants who reported a remote history of anthrax vaccination. Limitations to this data are the fact that antibiotics were begun before serum testing, and there were no baseline serum samples available for testing. Also, the time period from exposure to sampling was very short. Among 28 individuals with positive nasal swabs in the Capitol exposures who received antibiotics immediately, none had a positive culture from a nasal swab repeated 7 days later, and none had positive serum IgG to PA antigen 42 days after exposure. This again emphasizes the limitation and interpretation of a test in someone who had early antibiotic treatment. It does raise forensic considerations. Even with this easily disseminated strain, an antibody response may be aborted or modified with antibiotics by early eradication. Furthermore, antibiotics taken prior to exposure would likely be effective in preventing laboratory and clinical signs of an infection or exposure. Detection of DNA, antigen, or the organism itself on a person's body, clothing, or possessions would raise a red flag.

The route of infection is equally important in interpretation of results and the limitations of the assay used. The example of cutaneous anthrax in Paraguay illustrates this, as well as the notion of searching for other antigens as markers of exposure.[13] Analysis of an outbreak of at least 21 cases of cutaneous anthrax developed from touching raw meat of a sick cow was performed. Serum from 12 cases and 16 colony and two noncolony controls 6 weeks after the outbreak were blotted for antibodies to the PA and lethal factor components of anthrax toxin. An ELISA was used to probe for antibodies to poly-D-glutamic acid capsule. Of 12 cases, 11 had a positive PA screen, for a sensitivity of 91.7%; none of the 18 controls was positive for a specificity of 100. Only six of 12 cases had antibody to the lethal factor; all controls were negative. Probing for antibodies to capsule was positive in 11 of 12, but was positive in two of 18 controls. This study demonstrates the need to consider other antigens.

CONSIDERATIONS AND CONCERNS RAISED BY ANALYSIS OF OTHER INFECTIONS

Some of the principles discussed above are highlighted by a recent report on SARS. This coronavirus disease also evoked concern of a possible terrorist origin at the onset. A report in the Morbidity and Mortality Weekly Report (MMWR)[14] on the "Prevalence of IgG Antibody to SARS-Associated Coronavirus in Animal Traders" discussed the need to validate and interpret tests in the appropriate populations—the IgG test, discusses its inability to date the time of the infection, and the possibility of reactivity to a near neighbor that might be unknown. In a Promed bulletin, Dr. Berger looked at the data from a different angle and reported: "This week's study in MMWR indicates that animal contact may indeed promote infection; however, the most striking finding seems to have eluded the authors: 1.2 percent to 2.9 percent of individuals in a healthy control group of adults were also found to be seropositive! The population of Guangdong Province is 86.42 million (2001), of whom 61.14 million are adults over age 14. If we assume that the seropositivity rates among controls is representative of the province as a whole, 734,000 to 1,773,000 adults in Guangdong have at some time been infected by the SARS virus. These figures are 87- to 211-fold the total number (8,422) of SARS patients reported worldwide to date!" This is a good illustration of the need to question the methodology of acquisition of data before accepting their application in formulas or for analyses.

Yersinia pestis, the cause of plague, is a zoonotic infection which occurs in the U.S. with regularity and has an animal reservoir. This is in contrast to a case of smallpox which would raise an immediate red flag for a bioterrorist event. Cases need to be approached from an epidemiologic standpoint first to determine whether it is an "expected" case or whether the facts point to a deliberate introduction of the organism in a group of people or an individual. Analytic techniques could include genomic analysis of an isolated organism and immunological response of the host. In consideration of animal reservoirs, ELISA assays were compared with other tests for detection of plague antibody and antigen in multimammate mice (*Mastomys coucha* and *M. natalensis*).[15] They were experimentally infected and then sacrificed at daily intervals. IgG ELISA was equivalent in sensitivity to passive hemagglutination and more sensitive than the IgM ELISA and complement fixation. Antibody was detectable by day 6 after infection using all four tests. IgM ELISA titers fell to undetectable levels after 8 weeks. Plague fraction 1 antigen was detected in 16 of 34 bacteremic sera from *M. coucha* and *M. natalensis*. This shows that the principle of IgM versus IgG to this pathogen works to temporally situate the infection as early versus late or past. It also shows that when the information is combined with antigen detection, it engenders more confidence in the results.

Melioidosis is caused by *Burkholderia pseudomallei*.[16] It is also an example in which key clinical signs and laboratory features raise the possibility of this infection. Related studies and observations are presented here to illustrate some of the temporal issues of the host response and the need to interpret results of an assay in the appropriate clinical and geographic setting. Whether it is an acute infection, persisting one, or past infection can be determined by looking at several host responses. Often a simple indicator of infection such as erythrocyte sedimentation rate or C-reactive protein (CRP) can create clinical suspicion to begin a probe for a specific infection. In a study of 46 patients with clinical melioidosis, 35 (22 culture-positive and 13 culture-negative) had relatively uneventful disease courses. Initially they had elevated serum CRP that decreased with antibiotic therapy and returned to normal as their disease resolved. In another series of patients, IgM and IgG were measured by ELISA in 95 sera from 66 septicemic cases and 47 sera from 20 cases with localized melioidosis.[17] Sixty-five sera from culture-negative cases seronegative for other endemic infections but suspected of melioidosis were also examined. Other controls included 260 non-melioidosis cases, 169 high-exposure-risk cases, and 48 healthy individuals. The IgG-ELISA was 96% sensitive and 94% specific. All sera from cases with septicemic and localized infections and 61 of 63 sera from clinically suspected melioidosis cases were positive for IgG antibody. The sensitivity and specificity of IgM ELISA were 74% and 99%, respectively. A geometric index for IgM antibody in the sera of the melioidosis cases was significantly higher in melioidosis cases compared to that of the non-melioidosis disease controls. Another study by some of the same authors using a rapid test also showed IgG and IgM to have clinical utility.[18] Another study with the intent of evaluating the utility of an IgG assay compared to other assays illustrates how the clinical and temporal context must be integrated for interpretation.[19] It also illustrates how there is room for improvement in tests and that the best analysis will result from an understanding of the conditions in the endemic area and utilization of samples and controls from that area. These tests were evaluated in the actual clinical setting in an area endemic for melioidosis. Specificity of specific IgG (82.5%) and specific IgM (81.8%) were significantly better than that of an indirect hemagglutination test (74.7%). The sensitivity of the specific IgG assay (85.7%) was higher than that of the IHA test (71.0%) and the specific IgM test (63.5%). Specific IgG was found in septicemic cases (87.8%) and localized forms (82.6%). The specific IgG test was also better than the specific IgM test and the IHA test in identifying acute melioidosis cases in the first five days after admission. IgG antibody to a *B. pseudomallei* antigen remained high for longer than 5 years in recovered disease-free patients. Because this is a disease that may have an incubation of days to years, an acute case may very well be picked up by IgM versus IgG if it were a matter of days from infection. Although endemic for Southeast Asia, if it were used as a biothreat agent in a different environment, its etiology may not be recognized immediately. The importance of understanding endemic area factors as well as the host to the microbe is further illustrated in another zoonotic example. Rift Valley fever virus (RVFV) can be transmitted via aerosols. One study with the intent at looking for improved tests did show the utility of IgM to determine an early exposure to RVFV.[20] Two ELISA IgM tests detected specific IgM antibodies to RVFV during the first 6 weeks after vaccination. Three inactivated vaccine doses were given on days 0, 6 to 8, and 32 to 34. ELISA serum IgM on days 6 to 8 were negative or in the lower range of significance; on days 32 to 34 they were strongly positive; on days 42 to 52 they were waning and later were negative. The plaque reduction neutralization test was negative on days 6 to 8 and became positive in later samples. Similar to the examples shown above, their data suggest that three doses of RVFV vaccine induced a prolonged primary antibody response. The authors of that study concluded that the ELISA IgM could become an important tool for early diagnosis in acute human infection. Good correlation of a neutralization test and ELISA IgG would indicate a later infection.

Taken together these examples illustrate that an ideal test for both clinical and forensics use would incorporate endemic area controls, historical contextual information, knowledge of the route of exposure, background incidence, and kinetics of transmission.

POSSIBLE SCENARIOS OF BIOTERRORISM ATTACKS: DISTINGUISHING VICTIMS FROM PERPETRATORS

Each of these scenarios must take into account multiple factors and the limitations of any analytic process to be applied. This is often considered by understanding the elements of positive and negative predictive values of an assay within a population being tested. On one extreme is the situation that occurred with the onset of human immune deficiency (HIV) in the U.S. First there were no cases, and therefore a precise highly sensitive and specific test with excellent positive and negative predictive values (such as exists now when a combination of tests are used) would not likely yield a positive result in an area where there was little disease at the onset, Kansas, for example. A positive test by today's methodologies from a 1970 serum sample from Kansas would be considered a probable false-positive and warrant further investigation. However the same sample tested at the beginning of HIV testing could have been positive if the person had adult T-cell leukemia, which is caused by human T-cell leukemia virus-1 (HTLV-1). This is because the original tests for what became known as acquired immunodeficiency syndrome (AIDS) involved whole viral lysates in which up to 30% of the HTLV-1 sera cross-reacted. Suspicion to the contrary would be raised by knowledge of different presentations of the infection. For example, HTLV-1 can actually be used in the laboratory to immortalize cells. In the patient it actually increases the T-cell count, as is the nature with leukemia, instead of decreasing them, as with HIV infections. Other laboratory indicators such as hypercalcemia would now raise leukemia as a consideration.

Interpretation of a positive clinical test must take into account the health status of the person being tested. This is important for the practice of medicine and can have relevance when extended to forensic analysis. The following situations illustrate the concept. Individuals who have syphilis, a bacterial spirochetal infection, can typically have a positive FTA (fluorescent treponemal antibody) test for years. However while infected they would have a positive venereal disease research laboratory (VDRL) test. This reverts to negative with successful antibiotic therapy. There are some notable exceptions related to cross-reactive epitopes or autoimmune diseases. These are readily distinguishable by history and clinical information. Similarly individuals infected with tuberculosis will have a positive skin test (Mantoux), whereas the uninfected healthy person will be negative. In certain instances, a sick person with a cell-mediated immune deficiency will be anergic, that is, he/she will be negative to multiple skin tests including common antigens such as Candida. The key difference here is that there is a wide difference between the healthy person being tested and a very ill individual being subjected to the same test.

Tests may also discriminate between the time of the infection as acute or chronic, and its limitations may lead to different interpretations unless one is familiar with those limitations. An example of this occurred with the bacterial infection of *Borrelia burgdorferi*, which causes Lyme disease. Dattwyler's group showed that antibiotics could abrogate the antibody response because ELISA results were negative in 30% of patients with known disease who were treated early.[21] Another group showed that in early cases reactivity to a unique antigen, OspA, was also negative in serological assays despite a demonstrable T-cell response.[22] Our own group had an opportunity to analyze the same sera and found that there was antibody to *B. burgdorferi* but it was below the threshold of the conventional assays. It was detectable in its bound form, in immune complexes.[23,24]

Anthrax can be used as an example where investigatory leads can be generated by considering a scenario *in toto*. The elderly lady who died in Connecticut from anthrax clearly had no occupational exposure nor was she known to have had contact with anyone who had anthrax. It was possible that she received contaminated mail. However if this case had occurred as the index case or out of context of the mail attacks, it would have been reasonable to question her travel history, what her work if any was, or if she received or used products from an endemic area for anthrax. Similarly the Vietnamese woman who died in New York City would also have had these questions investigated. It would have been useful to search for direct or indirect evidence of anthrax by physical examinations of her contacts or close neighbors. Inspection and cultures from her workplace, apartment, and apartment complex (especially contiguous neighbors) are important for presence of anthrax. Coworkers, friends, neighbors, and other contacts could have had blood samples analyzed for antibody to anthrax antigens. These samples could have been frozen so that if one were positive it would be available for a comparison study in the future. At a minimum these types of studies could serve as future control data for the geographic region. Although hypothetical, several results could have occurred, and each will be analyzed separately: First example: a close contact is positive for IgM to one of the *B. anthracis* antigens, e.g. PA. This would suggest that this person had recent exposure and if nothing else should be treated. This individual could conceivably be the one who knowingly or unknowingly passed the spores to the patient. Given the October 26 onset of illness, which is late in the mailing sequence, it would be less likely that this individual was a perpetrator but rather a recent victim too. However if this person were IgG-positive on the assay, then there are several other possibilities. Perhaps this person had past exposure in an endemic region with a subclinical or treated illness (e.g., Haiti, where anthrax is known as "charcoal disease"). Or this person could have been vaccinated for bona fide reasons such as a researcher who received it for occupational protection. Or this person could have obtained the vaccine originally for legitimate or illegal purposes but was nevertheless vaccinated. Animal vaccines may be more obtainable without strict record keeping. This person could have loaded the mail with relative impunity if there were protective antibody generated from the vaccine. These situations require intelligence information regarding access, ability, and motive. However the IgG finding could point investigators towards such an individual, whereas an IgM finding justifies critical therapy. Coming from the other direction, where information points to a particular individual, investigation could be extended to ingestion of antibiotics. Questions would be raised regarding access to antibiotics, recent ingestion of them, half-life of the antibiotic, half-life of the metabolites of the antibiotics, and in which body fluids or tissues the residual can be found. As illustrated from the data in the earlier sections, someone with antibiotics in their system could be protected from exposure to a sensitive microbe. This person would be antibody-negative and likely antigen- and microbial DNA/RNA-negative, since the infection would have been eradicated before the organism could proliferate in any significant quantity.

Similar strategies can be employed to examine suspicious but possible accidental transmission of infections. This is illustrated by a recent series of aviant flu. Tools to determine a person to person spread as the transmission mode included viral cultures, serologic analysis, immunohistochemical assay, reverse-transcriptase-polymerase-chain-reaction (RT-PCR) analysis, and genetic sequencing.[25]

It is likely that future understanding of the immune system and evolving technologies such as microarrays will bring new analytic power to the scene, but in the meantime we can make good use of proven principles for forensic purposes.

REFERENCES

1. Budowle B, Schutzer SE, Einseln A, Kelley LC, Walsh AC, Smith JA, Marrone BL, Robertson J, Campos J. Public health. Building microbial forensics as a response to bioterrorism. *Science* 2003;301:1852–1853.
2. United States, Federal Bureau of Investigation, Laboratory FBI. Handbook of forensic services. 1999.
3. Averill DC, American Society of Forensic Odontology. Manual of forensic odontology. S.l.: American Society of Forensic Odontology: David C. Averill; 1991.
4. Janeway C. *Immunobiology 5, The immune system in health and disease*. New York: Garland Publishing; 2001.
5. Paul WE. *Fundamental Immunology*. New York: Raven Press; 1993.
6. Parslow TG. *Medical Immunology*. New York: Lange Medical Books/McGraw-Hill Medical Pub. Division; 2001.
7. Mandell GL, Douglas RG, Bennett JE, Dolin R. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*. Philadelphia: Churchill Livingstone; 2000.
8. Jackson PJ, Hugh-Jones ME, Adair DM, Green G, Hill KK, Kuske CR, Grinberg LM, Abramova FA, Keim P. PCR analysis of tissue samples from the 1979 Sverdlovsk anthrax victims: The presence of multiple Bacillus anthracis strains in different victims. *Proc Natl Acad Sci U S A* 1998;95:1224–1229.
9. Traeger MS, Wiersma ST, Rosenstein NE, Malecki JM, Shepard CW, Raghunathan PL, et al. First case of bioterrorism-related inhalational anthrax in the United States, Palm Beach County, Florida, 2001. *Emerg Infect Dis* 2002;8:1029–1034.
10. Jernigan JA, Stephens DS, Ashford DA, Omenaca C, Topiel MS, Galbraith M, et al. Bioterrorism-related inhalational anthrax: The first 10 cases reported in the United States. *Emerg Infect Dis* 2001;7:933–944.

11. Quinn CP, Semenova VA, Elie CM, Romero-Steiner S, Greene C, Li H, et al. Specific, sensitive, and quantitative enzyme-linked immunosorbent assay for human immunoglobulin G antibodies to anthrax toxin protective antigen. *Emerg Infect Dis* 2002;8:1103–1110.
12. Dewan PK, Fry AM, Laserson K, Tierney BC, Quinn CP, Hayslett JA, et al. Inhalational anthrax outbreak among postal workers, Washington, D.C., 2001. *Emerg Infect Dis* 2002;8:1066–1072.
13. Harrison LH, Ezzell JW, Abshire TG, Kidd S, Kaufmann AF. Evaluation of serologic tests for diagnosis of anthrax after an outbreak of cutaneous anthrax in Paraguay. *J Infect Dis* 1989; 160:706–710.
14. Centers for Disease Control and Prevention. Prevalence of IgG antibody to SARS-associated coronavirus in animal traders. *MMWR* 2003;52:986–987.
15. Shepherd AJ, Hummitzsch DE, Leman PA, Swanepoel R, Searle LA. Comparative tests for detection of plague antigen and antibody in experimentally infected wild rodents. *J Clin Microbiol* 1986;24:1075–1078.
16. Ashdown LR. Serial serum C-reactive protein levels as an aid to the management of melioidosis. *Am J Trop Med Hyg* 1992;46:151–157.
17. Chenthamarakshan V, Vadivelu J, Puthucheary SD. Detection of immunoglobulins M and G using culture filtrate antigen of Burkholderia pseudomallei. *Diagn Microbiol Infect Dis* 2001;39:1–7.
18. Cuzzubbo AJ, Chenthamarakshan V, Vadivelu J, Puthucheary SD, Rowland D, Devine PL. Evaluation of a new commercially available immunoglobulin M and immunoglobulin G immunochromatographic test for diagnosis of melioidosis infection. *J Clin Microbiol* 2000; 38:1670–1671.
19. Dharakul T, Songsivilai S, Anuntagool N, Chaowagul W, Wongbunnate S, Intachote P, Sirisinha S. Diagnostic value of an antibody enzyme-linked immunosorbent assay using affinity-purified antigen in an area endemic for melioidosis. *Am J Trop Med Hyg* 1997; 56:418–423.
20. Niklasson B, Peters CJ, Grandien M, Wood O. Detection of human immunoglobulins G and M antibodies to Rift Valley fever virus by enzyme-linked immunosorbent assay. *J Clin Microbiol* 1984;19:225–229.
21. Dattwyler RJ, Volkman DJ, Luft BJ, Halperin JJ, Thomas J, Golightly MG. Seronegative Lyme disease. Dissociation of specific T- and B-lymphocyte responses to Borrelia burgdorferi. *N Engl J Med* 1988;319:1441–1446.
22. Krause A, Burmester GR, Rensing A, Schoerner C, Schaible UE, Simon MM, Herzer P, Kramer MD, Wallich R. Cellular immune reactivity to recombinant OspA and flagellin from Borrelia burgdorferi in patients with Lyme borreliosis. Complexity of humoral and cellular immune responses. *J Clin Invest* 1992;90:1077–1084.
23. Schutzer SE, Coyle PK, Dunn JJ, Luft BJ, Brunner M. Early and specific antibody response to OspA in Lyme Disease. *J Clin Invest* 1994;94:454–457.
24. Schutzer SE, Coyle PK, Belman AL, Golightly MG, Drulle J. Sequestration of antibody to Borrelia burgdorferi in immune complexes in seronegative Lyme disease. *Lancet* 1990; 335:312–315.
25. Ungchusak K, Auewarakul P, Dowell SF, Kitphati R, Auwanit W, Puthavathana P, Uiprasertkul M, Boonnak K, Pittayawonganon C, Cox NJ, Zaki SR, Thawatsupha P, Chittaganpitch M, Khontong R, Simmerman JM, Chunsutthiwat S. Probable person-to-person transmission of avian influenza A (H5N1). *N Engl J Med* 2005;352:333–340.

CHAPTER 15

Bioinformatics Methods for Microbial Detection and Forensic Diagnostic Design

TOM R. SLEZAK

Lawrence Livermore National Laboratory, Livermore, California

STEVEN L. SALZBERG

Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland; The Institute for Genomic Research, Rockville, Maryland

This chapter reviews bioinformatics techniques that can be used for detection and identification of microbial samples, and for further analysis of sequence information derived from microbial organisms. We begin with an overview of laboratory techniques used to create the diagnostic and forensic data, and then describe methods for sequencing, assembling, and analyzing genomic data. We then consider algorithms for large-scale genome alignment and comparison, followed by methods for modeling protein structure. The chapter concludes with real-life examples demonstrating working systems for detection and analysis of the microbial pathogens, using both DNA and protein detection methods.

A WORKING DEFINITION OF BIOINFORMATICS

The term *bioinformatics* entered the lexicon during the early years of the Human Genome Project as it became necessary to apply a wide range of computational techniques that exceeded what the average bench biologist was capable of accomplishing on a desktop computer. An extremely elastic term, bioinformatics has been stretched to cover anything remotely computational that is applied to anything remotely connected to modern biology. For the

Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

purposes of this chapter, we will confine the discussion of bioinformatics to topics that are directly connected to our perception of computational approaches for microbial detection, diagnostics, and forensics. Due to length restrictions we will undoubtedly not fully cover some bioinformatics topics that could reasonably be tied to microbial forensics.

AN OVERVIEW OF MICROBIAL DIAGNOSTICS

The goal of bioinformatics in microbial forensics and diagnostics is to support the rapid identification of microbial samples at a variety of levels of resolution (family, species, strain, isolate). Techniques to accomplish this are ultimately based upon accurate determination and comparison of genomic sequence of high quality, combined with detailed analysis of the sequence to determine gene composition, protein structure, and the functional organization of the mechanisms of transmission and virulence. As a prelude to the bioinformatics discussion and as motivation for some of the techniques, we will first introduce some of the laboratory techniques involved in modern microbial forensics.

A broad range of casually defined terms has evolved to describe the processes used to diagnose the presence of bacterial or viral agents. Microbiologists have used *culture-based diagnostics* for decades to grow and characterize bacteria that are capable of being grown *in vitro*. Prior to the need to be concerned with bioterror attacks, the public health system was concerned with *clinical diagnostics* that could distinguish between the wide variety of naturally occurring pathogens that present similar symptoms. For a number of years these have included both classical culture-based diagnostics as well as modern *genomics-based diagnostics*, which includes the more specific category of *polymerase chain reaction* (*PCR*)*-based diagnostics*. An important concept regardless of technique is *rule-out diagnostics*, which eliminate the pathogen in question and may even confirm the presence of a microbe of less concern (e.g., chickenpox instead of smallpox). The same can be said for microbial diagnostics for the agriculture and food processing sectors.

Subsequent to the 2001 anthrax attacks, increased distinctions have been drawn between *detection diagnostics*, which attempt to indicate the presence of specific pathogens at a broad level (family, genus, or species), and *forensic diagnostics* that resolve individual strains or isolates. Our discussion will focus on detection and forensic diagnostics from the viewpoint of detecting and analyzing bioterrorism incidents. While these techniques have obvious additional applications in clinical health, agriculture, and food safety, we will not address these areas in detail.

Note that detection and forensic diagnostics are focused on answering the question of which organisms are present in a given sample. The possibility remains that through extensive genetic engineering, *identity-based diagnostics* could be successfully spoofed. An alternate approach is to perform *function-based diagnostics* that would focus on what potential damage the sample could do to a human (or plant or animal) host. Although in its relative infancy, this approach would focus on minimum gene “kits” or cassettes that could transfer virulence to an otherwise benign host microbe.

It is useful to examine the impact of economics on the use of microbial diagnostics. Detection diagnostics are employed to provide wide-area protection against the possibility of a pathogen release, in the case of bioterrorism, or to monitor the spread of a natural pathogen (e.g., West Nile virus). Depending on the circumstances of the surveillance and the probability of encountering the pathogen, species-level diagnostics (“is there any anthrax/plague/West Nile virus present in this sample?”) are usually appropriate. However, if the probability of encountering an individual species is extremely low (e.g., smallpox), it may make economic sense to use a broader-range diagnostic (e.g., orthopox-wide).

Forensic diagnostics are employed when detection diagnostics have confirmed that one or more pathogens of interest are present in a sample and it is desired to obtain a finer level of resolution about which strain or isolate is present. In general, forensic diagnostics currently are much more expensive than detection diagnostics. Pathogen detection may someday reach a point where all known strains of all known pathogens can be detected in parallel on a low-cost, high-density array or “chip” or future technique, but until then it is likely that a hierarchical approach to detection and forensics will be employed.

DETECTION DIAGNOSTICS

Detection diagnostics may be divided into two major categories based on issues of speed, cost, and sensitivity. *Nucleic acid diagnostics* are used when high sensitivity is required. DNA extraction and amplification steps add extra time and reagent costs, but provide sensitivity down to the single-copy level. *Protein diagnostics* are used when low cost and fast results are desired and low sensitivity can be tolerated (e.g., many thousands of copies can be assumed to be present in the sample). Most protein diagnostics are focused on detecting proteins expressed on the surface of the pathogen in one or more states of its life cycle. This avoids extra steps to disrupt the pathogen and expose internal proteins; however, this approach is always available if required. Other protein detection diagnostics are aimed at detecting products that may be released by the cell. Home pregnancy tests are a common example of protein detection technology.

Not surprisingly, the two approaches are complementary and can be used in tandem, with protein diagnostics acting as an event trigger, with confirmation followed up by nucleic acid diagnostics. The high mutation rates of some RNA viruses results in insufficient nucleotide conservation for species-wide nucleic acid diagnostics. Most of this variation is in the 3rd base position of the codon (the "wobble base"), meaning that sufficient protein sequence conservation remains to allow for species-level protein detection. It is worth noting that any specific nucleic acid signature could be similarly spoofed (i.e., via genetic engineering) by changing 3rd positions of codons in a way that did not change the resulting protein sequence. Protein-based detection would be one defense against this level of sophistication of genetic engineering. It is a safe bet that a range of both nucleic acid and protein detection techniques will always be useful for pathogen diagnostics. It should also be noted that pathogens might be detected indirectly via *host response* effects that could lead to different expression levels or the presence of proteins that would not be seen in unaffected individuals. Diagnostics based on host pathogen response are in their infancy, and could be used either for detection or forensic diagnostics.

NUCLEIC ACID DETECTION DIAGNOSTICS

Nucleic acid detection diagnostics are based upon the concept of conservation of sufficient nucleotides of genomic sequence across all strains to allow their detection via one or at most very few assays. This can either be determined via sequencing all known strains, or assumed after sequencing a sample of them and testing the resulting diagnostics against the full strain panel. Note that besides being *conserved* across all strains, the nucleic acid diagnostics (often called *signatures*) must also be *unique* (i.e., they do not cross-react with any other organisms). This implies that all nucleic acid diagnostics run the risk of eventual false-positive failure, should they be tested against a sample that contains an organism with the same signature DNA. It also implies that all nucleic acid diagnostics run a risk of eventual false negative failure, should a strain exist or newly evolve that does not maintain conservation in the chosen signature region(s). Together, these imply that nucleic acid signatures need to be frequently checked for continued apparent conservation and uniqueness as new sequence information is obtained.

As we will see in the section on basecalling accuracy below, *forensic diagnostics* for nucleic acids can be remarkably accurate. By using validated statistics from individual DNA sequences combined with redundant testing, we can increase our confidence in the identity of a particular nucleotide until the chance of error is vanishingly small. Thus even if two isolates of a given bacterium differ in only a single nucleotide, we can still determine with extremely high confidence that they are different. The techniques to be described below illustrate how to first identify such tiny differences and then how to evaluate their probability of error.

CHEMISTRIES FOR NUCLEIC ACID DETECTION

A variety of chemistries are available for implementing nucleic acid detection via PCR. Differences occur in technical details of primers and probes, licensing requirements, reagents, design software, and costs. TaqMan® PCR, a technique in common use for nucleic acid detection, involves a two-step process. The first step uses a pair of PCR primers to amplify the target region. The second step uses a *hybridization probe* with a fluorescent tag to locate a unique region on the amplicon. Detection and measurement of the fluorescence provides information about the presence and quantity of the target in the original sample. Detection assays based on TaqMan® PCR have been successfully used to detect natural pathogen outbreaks (http://www.llnl.gov/str/JulAug01.html) and are in regular use protecting a number of U.S. cities.

Molecular Beacons (http://www.molecular-beacons.org/) are another nucleic acid diagnostic method. These are oligonucleotide probes that become fluorescent upon hybridization with target DNA or RNA molecules. Biological detection based on molecular beacons does not need labeled analyte or intercalation reagents and is direct, sensitive, and highly selective. Probe immobilization, which is essential for conventional DNA hybridization, becomes unnecessary and thus the protocol for optical detection becomes simpler. Lower per-test cost and less expensive licensing are reasons for a surge of recent interest in molecular beacons.

SINGLE AND MULTIPLEXED FORMATS

The cost of nucleic acid diagnostics can potentially be greatly reduced if multiple reactions can be processed simultaneously in the same tube or well. While simple in concept, successful deep multiplexing of four or more singleplex assays can be extremely difficult. The major difficulty is that assays that perform perfectly well on their own can have major cross-reactions when combined. Vendor-specific software to design multiplex assays is available; the depth of multiplexing possible may vary and some software may be better at checking potential multiplexed assays than at generating them.

PLATFORMS

A wide variety of commercial instruments are available to perform nucleic acid detection diagnostics. They deliver a range of features and sensitivities, and have different total operating costs based on throughput, monopoly on reagents, and other factors. Each may be best suited to a particular work environment, duty cycle, or other parameter.

The handheld advanced nucleic acid analyzer (HANAA), the first miniaturized PCR unit, was invented at Lawrence Livermore National Laboratory (LLNL) (http://www.llnl.gov/str/JanFeb02/pdfs/01_02.4.pdf). The technology was subsequently licensed to industry, and now a range of systems based on several technologies are available from Cepheid (http://www.cepheid.com), Idaho Technologies (http://www.idahotech.com), Applied Biosystems (http://www.appliedbiosystems.com), Bio-Rad (http://www.bio-rad.com/iCycler), and others.

Significant differences exist in design philosophies and capacities. Cepheid currently uses 16 independently-programmable chambers, each taking a single reaction tube, and is capable of delivering results in 30 minutes or less. Applied Biosystems uses 96-well plates (or equivalent format tubes) and produces results in about 2 hours. Significantly, although all instruments have good graphical user interfaces suitable for use in research labs, none yet appears to have a good programmatic interface that allows automatic programming and execution of an analysis, using a sample sheet delivered over the network, and seamless delivery of easily parsed results to an external controlling system. These deficiencies make it difficult and potentially error-prone to use these devices in large-scale, continuous pathogen detection systems.

Luminex (http://www.luminexcorp.com) has tackled the need to increase analysis throughput using a novel approach. Currently, a set of 100 different colors of microsphere beads is available that can be differentiated with a combination of two lasers. Each of the 100 bead types can be dedicated to a single nucleic acid diagnostic. A third laser color can be used to detect the hybridization reaction when the beads are passed through the illumination chamber via a flow-sorting apparatus. The system is also capable of being used for protein-based detection, unlike other current instruments. It may be possible to increase the bead resolution greatly, providing forensic-level capabilities as well.

PROTEIN DETECTION DIAGNOSTICS

Protein detection diagnostics run similar risks for the assumptions of conservation within the species and uniqueness compared to all other organisms, as described for nucleic acid diagnostics above. One advantage of protein detection is that it allows the toxins created by certain pathogens to be detected in circumstances where the organism itself is no longer capable of being detected via nucleic acid techniques (e.g., *botulinum* toxin). It also is harder to engineer around protein-based detection and still retain the original virulent characteristics of the pathogen.

Two types of protein diagnostics are currently in use. Antibodies are complex proteins that bind to an exposed region of a target protein. Monoclonal antibodies, which are derived from cells called "hybridomas," are preferred over polyclonal antibodies in that the cell line producing the monoclonal antibody is immortalized. It is therefore capable of producing a virtually unlimited supply of the antibody. Monoclonal antibodies are also preferred because they recognize a single epitope (binding site) on the target molecule. High-affinity ligands (HALs) are small molecules that are selected to fit and bind uniquely into crevices or pockets in the target protein surface. In either case, the diagnostic agent may be linked to a reporter molecule (e.g., biotin, fluorescent tag) that can be easily detected. In the case of antibody detection, it is more common to use a tagged secondary antibody to detect the bound monoclonal. Recall that most protein detection assays do not include an amplification step, thus significantly reducing the sensitivity of this approach compared to PCR-based techniques.

MONOCLONAL ANTIBODY

Monoclonal antibodies are produced *in vivo* by injection of a target protein (the "antigen") into mice. However, the outcome of immunization and subsequent antibody production depends largely on complex biological conditions. An alternate approach, called "phage display," offers clear advantages, such as cost and speed, and suggests the further capability of genetically engineering specific peptides with high affinity to any target protein. To date, attention has focused on selecting short, unique, surface-accessible peptides on the target protein for injection into mice. However, it is conceivable that antibodies could be produced that recognize closely spaced but discontinuous peptides or surface-exposed residues contained within protein secondary structural units, such as alpha-helices. We are not aware of any publicly available software that can automate the selection of suitable discontinuous peptides or residues on the target protein for improved antibody production, or that can assist in the engineering of protein-based ligands.

HIGH-AFFINITY LIGANDS

Production of HALs currently requires that the protein structure be known experimentally. The surface is manually inspected by experts using programs such as *DOCK* (http://www.cmpharm.ucsf.edu/kuntz/dock.html) to identify suitable-sized surface depressions or pockets. A large library of synthetic molecules is then compared against the pocket shape, looking for a potential "fit." Potential fits are further examined for features such as charge, hydrophobicity, etc. that could preclude them from successful binding. Synthetic chemistry is used to create the best candidates, which are then tested against the target protein. This process has been used to create HALs that recognize various pathogen toxins, as seen in Figure 15.1.

At LLNL we are exploring several ways to scale up this process. First, as described below, we have improved protein structure modeling so that HAL design can be attempted without an experimentally solved structure. Second, we are investigating faster methods for locating suitable pockets. Finally, we plan to incorporate large supercomputers to speed up checking of the small

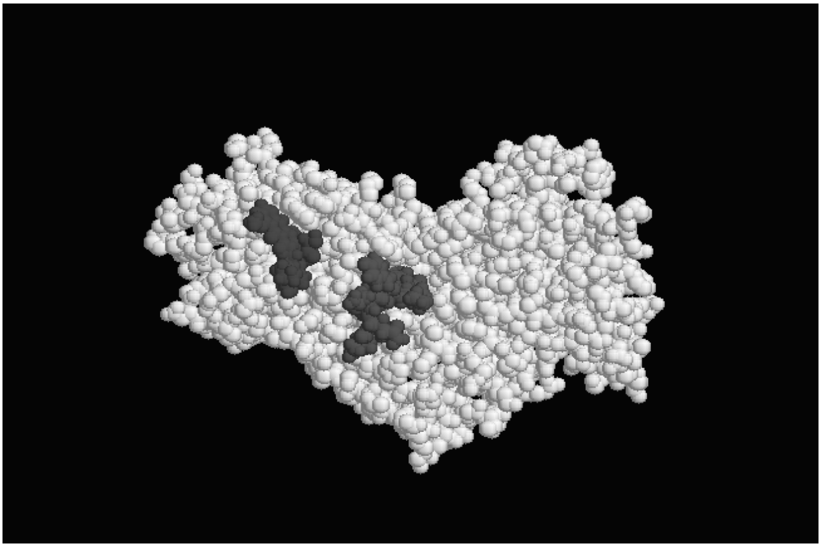


FIGURE 15.1 A representation of a bi-dentate ligand that recognizes the botulinum toxin. Figure courtesy of Rod Balhorn, LLNL. (See color insert.)

molecule libraries against each potential HAL pocket. Other similar efforts are underway, and it is hoped that a major acceleration of protein detection diagnostics can be achieved within a few years.

FORENSIC DIAGNOSTICS

Forensic diagnostics are primarily characterized by one or more assays (often run in parallel) that can establish with a high degree of certainty a strain or isolate level of resolution. Some forensic assays are unique (e.g., one assay per strain/isolate) while others establish a *fingerprint-like* pattern (e.g., run *k* assays that deliver binary or higher-order results, the pattern of which can be used to establish strain/isolate resolution). At this time, nucleic acid forensic diagnostics are somewhat more mature than protein forensic diagnostics, reflecting the early stages of proteomic knowledge and application.

High-quality data from multiple strains/isolates are required to establish good nucleic acid forensic diagnostics. Some techniques (e.g., RFLP, see below) do not require genomic sequencing of each strain and may prove adequate for strain differentiation for some organisms. Many organisms require a finer granularity of variation to distinguish strains and thus require rather high-quality sequencing.

Restriction fragment length polymorphisms (RFLPs) are a long-established genomic fingerprint-like technique capable of providing at least a moderate degree of resolution between strains/isolates of bacterial pathogens. One or more restriction enzymes are used to digest a pathogen genome; resulting fragment lengths are determined from a gel and can then be compared with a reference known-strain fingerprint library. This technique is the basis of "DNA fingerprinting" used by modern law enforcement in recent years (http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/R/RFLPs.html).

Researchers have noted that bacterial strains often contain variation in simple nucleotide repeat patterns. Presumably caused by DNA replication errors over evolutionary time, they can be exploited by a variety of techniques to provide strain- or isolate-level resolution. Variable number tandem repeats (VNTRs) are short (typically 1–20-bp) sequences that occur in different copy numbers from one strain of a bacterium to another. By identifying the number of copies of each known VNTR, strains can be identified and distinguished from one another. The same principal has been applied to the human genome, where similar short, variable repeats have been given the name "microsatellites." Once a VNTR has been identified, it can be assayed using PCR primers from either side of the repeat region. After PCR amplification, the size of the resulting PCR product can be measured precisely on an agarose gel, and the number of copies can be determined from the size. If the repeat region is very small, e.g., one nucleotide, then this size assay may not be sufficiently accurate. In such cases it may be necessary to sequence the PCR product to validate the estimate made from the size assay.

By typing multiple VNTRs simultaneously, one can create a very sensitive genetic signature of a strain or isolate, even for species that show relatively little between-strain variation. This technique, multiple-locus VNTR analysis (MLVA),[1] has allowed researchers to track the spread of plague isolates across a set of prairie dog colonies.[2] As above, PCR primers are developed to extract the VNTR regions. The resulting fingerprint of the presence and length of each VNTR provides higher resolution than a single VNTR in isolation, which can be especially important for genotyping samples that have experienced many changes since divergence from known samples. In such cases, where many of the VNTR loci may differ from each previously characterized isolate, a set of VNTRs may allow one to determine the most closely related strain with respect to a new sample.

In recent years, analysis of individual variation in humans and other species has identified large numbers of single nucleotide polymorphisms (SNPs) that can be analyzed to provide resolution down to the individual organism, in some cases. This technique can apply equally to bacterial and viral pathogens. It requires high-quality genomic sequence so that SNPs can be correctly distinguished from sequencing or assembly errors. It also requires careful analysis to establish a subset of all apparent SNPs that is capable of accurately resolving all of the known strains. Application of SNPs for forensics identification of an unknown sample requires accurate sequencing of the sample in at least all of the regions covering the SNP subset.

To determine the identity of a single SNP, one can use PCR or other sequence-based assays, but high-volume SNP detection is more effectively accomplished through hybridization methods, in which many short oligomers hybridizing to the region in question can be assayed simultaneously. This can be achieved through a chip-based assay, described below.

The larger task is that of creating the set of diagnostic SNPs (and other genetic markers) in the first place. The ideal method is to sequence the complete genomes of three or more strains chosen to represent the genetic diversity of the species. These can then be aligned (using methods described below) and all differences can be found, including VNTRs, SNPs, and larger insertions and deletions. This approach, although somewhat more costly initially, yields a valuable set of forensic tools, as it provides an unbiased sample of all genetic differences between a set of strains. Unlike earlier methods, which identified only a select subset of polymorphisms, whole-genome sequencing identifies everything. The first use of whole-genome sequencing for forensic analysis was conducted in response to the 2001 anthrax attacks, and it resulted in the discovery of 60 novel polymorphisms in *B. anthracis*.[3] And while it may seem relatively expensive, the scientific community already is approaching a goal of at least one complete genome for all major human pathogenic bacteria and viruses, and we have two or more strains for many others. The advantage to this approach for diagnostics is that once the genomes are sequenced, they become a valuable public resource that can be used by countless future investigators, not only for forensics but also for research on clinical diagnosis and treatment. The Institute of Genomic Research (TIGR) is currently sequencing 15 or more strains of *B. anthracis* with the express purpose of creating a comprehensive set of genetic markers for that species.

When using whole-genome sequencing in a high-throughput manner with the purpose of identifying polymorphisms, it may be most effective to produce a "draft" sequence, in which the random shotgun phase proceeds until approximately 99% of the genome is covered. This leaves many gaps and other regions with low coverage, and therefore low accuracy. As a result, apparent polymorphisms found in such regions may be false positives. We discuss methods for calculating precise probabilities for each such polymorphism in the section on genome annotation for diagnostics, below.

LARGE-SCALE, CHIP-BASED TECHNIQUES

Techniques such as VNTR, MLVA, and SNP can be implemented into microbial forensic diagnostics via sets of independent assays, each focused on a single VNTR or SNP. The results of these multiple assays can then be combined to provide an identification verdict. The use of large scale oligonucleotide chips provides a massively parallel approach to providing strain/isolate-level forensic resolution for a large number of pathogens in a single assay. These chips can hold 500,000 or more (20 to 30 base) oligomers fixed to a glass substrate; application of the unknown sample thus provides a massive hybridization-event fingerprint. With this large number of oligomers, it is possible to redundantly target each base in the region of each polymorphism site (SNP or repeats) with both perfect match and perfect mismatch oligos. The ratio of hybridization intensities at each base location is used to determine the presence or absence of each oligomer in the sample.

Several projects underway have as their goal the production of a "pathogen chip" capable of providing accurate strain resolution for all major pathogens. Such a chip could also help detect engineered organisms, due to the parallel nature of the chip analysis. Drawbacks of this approach include the high cost of chip mask design, the current high per-use cost (and one-time use) of the chips, and the high cost of keeping the chip current as new strains are sequenced. It is also not easy to optimize hundreds of thousands of probes to all work under the same conditions, which leads to a certain degree of difficulty in interpreting the resulting data. To address some of these economic concerns, a number of competing technologies are in various stages of research. It is likely that the cost and effectiveness of highly parallel nucleic acid microbial forensic techniques will greatly improve within the decade.

PROTEIN FORENSIC DIAGNOSTICS

Traditionally, proteomics techniques are used to detect *relative expression levels* of an organism's proteome for a specific tissue sample and time state. For example, which genes of *Yersinia pestis* are up-regulated in a flea host but down-regulated in a mammalian host, or vice versa? Given sufficient viable sample, it may also be possible to use proteomic forensic techniques in some situations.

One can imagine having reference *microarrays* or *protein chips* containing the proteomes of all known strains of an important pathogen. In situations where other forensic techniques could not positively identify a sample, perhaps indicating an engineered pathogen, a decision could be made to run the sample against the proteome microarrays. This might rapidly indicate situations where more-virulent genes from one strain were engineered into a less-virulent one. This approach, used properly, might give results far quicker and cheaper than sequencing the sample.

Also, as mentioned above, bead-based technologies like that currently available from Luminex are also capable of providing protein-based detection at a density that makes them suitable for consideration for protein forensic diagnostics.

GENOME SEQUENCING AND ANALYSIS TECHNIQUES

Pathogen sequencing is the basis for establishing virtually all genomics-based microbial detection and forensic diagnostics. A range of strategies exists for obtaining and exploiting DNA sequence for diagnostic signatures. It is worth noting that sequencing performed with a primary goal of aiding microbial forensics may be done with different economics than sequencing performed for more traditional academic reasons. For example, assembly of the K^{th} strain of a bacterial pathogen could take full advantage of the knowledge gained from assembling the prior strains, instead of proceeding with a *de novo* assembly. Deep-draft contigs could be layered upon one or more finished strains to facilitate rapid gap closing; in fact, this is exactly how the multi-strain *B. anthracis* project at TIGR was designed[3] and was the motivation for a new comparative genome assembler that assembles a draft onto a reference.[3a]

BASECALLING AND ACCURACY

For microbial forensics, we need to compare two genomes and discover all differences between them. Once an initial set of differences is found, we need to determine whether they are all correct. Obviously, if the sequences are correct, then every difference represents a distinguishing genetic marker. But genome sequences are not 100% correct, and the accuracy varies from one nucleotide to the next. The most important factor in calculating the accuracy of a base is *coverage*.

When a genome is sequenced using the whole-genome shotgun (WGS) method, which today is standard, each base is sequenced independently seven or eight times, on average.

The level of sequencing redundancy is called coverage, and it can be explained simply as follows. To prepare the DNA for sequencing, total genomic DNA is sheared randomly (using sonication or other methods) and then size-selected to produce a genomic library of a certain size, for example 3 kbp. The library consists of a large set of 3 kbp inserts in a clone vector such as pUC18 (many standard and customized vectors have been designed, with preferences for different insert sizes). Sequences are prepared from these clones, and both ends of each insert are sequenced. A sequencing "read" in 2005 contains ~700–800 nucleotides (longer reads can be obtained by using different reaction mixtures), a number that is steadily getting larger as technology improves. When we say a genome was sequenced to 8× coverage, we mean that each nucleotide in the genome is, on average, contained in eight separate reads. If the original clone library represents a perfectly uniform random sample of the genome (which it never does, but the approximation is useful), then 8× coverage implies that over 99% of the nucleotides in the genome are contained in at least one read.[4]

The length of a read is determined not only by the sequencing chemistry and technology, but also by the software that calls bases. The most commonly used basecalling programs are phred,[5] TraceTuner, and KB (commercial products designed for Applied Biosystems 3700 capillary sequencers). Each of these programs converts the four-color signal (the chromatogram) generated by an automated sequencer into a series of bases, each of which also has a quality value attached. These quality values are simply error probabilities converted to a more intuitive range, using the formula $Q = -10 \log P$, where P is the probability that the basecall is in error. Thus if the probability of error is 1/1,000, then $Q = 30$. Every major sequencing center, and every genome project, uses these programs or similar ones, and thus has a quality value attached to every base.

To understand the effect this has on the accuracy of assembled DNA sequences (see the next section for more on assembly), consider the possible scenarios where coverage ranges from 1× to 8×. Assuming that each read containing a particular genome position is independent (again, not a perfect assumption, but a reasonable approximation), the probability of making an error in a basecall can be calculated by multiplying the error probabilities in each of the reads. Thus at 1× coverage, a base reported with a quality value of 20 has a probability of 1% of being wrong. In contrast, if that same position has 8× coverage, and if each of the eight reads have a quality value of 30, and if all the reads agree on the identity of the base, then the probability of error is only 10^{-24}. (This calculation assumes that the probabilities of error for each base in each read are independent, an admittedly over-simplistic assumption.) At this level of accuracy, we would only expect to see no errors at all in any genome, even a large mammalian genome.

Thus using quality values and some straightforward assumptions, we can compute the accuracy of any base in a genome sequence. Using similar principals, we can compute the likelihood that a difference between two genomes is real. The statistics behind this computation were first explained in the study comparing the *B. anthracis* used in the October 2001 attacks to the reference Ames strain.[3] That study also showed how to extend the statistical model to compute a confidence in a VNTR. The idea is that if any of the "extra" bases in the longer of two VNTR regions are correct, then there is a genuine difference between the strains being compared.

Note that for forensic purposes, if an SNP or VNTR appears to have an unacceptably high probability of error (using the calculations described above and in Read et al.[3]), then one can reduce the probability by retesting. As long as the original sample material is still available, a basecall can be reconfirmed by resequencing the region in question. Each additional sequence produces a basecall along with a probability of error, and this probability can be multiplied by the previous value to produce a new, usually lower likelihood of error. (If the resequencing gives a different answer, though, then the likelihood of error will increase rather than decrease.)

The problem of basecalling accuracy is still an area of active research, and more precise estimates of accuracy may emerge as programs are refined further. For example, some recent SNP analyses have shown that if the quality values around a given nucleotide are uniformly high, then one can confidently revise a quality value upward.[6] Perhaps the most direct way to improve accuracy is to consider every SNP to be a potential miscalled base, and to reanalyze the raw chromatogram data from each underlying read. A new system, AutoEditor, that does exactly this was released in mid-2003, and was shown to correct 85% of the mis-called bases in a large collection of over 25 complete genomes.[7] This dramatically reduces the potential set of SNPs in a genome sequence, and allows scientists to focus more effectively on genuine polymorphisms.

DRAFT VERSUS FINISHED

Much of the time and cost of sequencing is involved in closing gaps and raising the entire genome to a specified level of quality (e.g., an error rate of less than 1 in 10,000 bases). This level of quality is assumed for traditional academic sequencing where full annotation of the organism's genes is desired, and is highly desirable for at least one strain/isolate for any pathogen. But is this level of quality, effort, and cost required for all additional strains/isolates of a pathogen? The answer depends in large part on the resolution of diagnostics being designed, the nature of the pathogen, and available time and funding.

For detection diagnostics it is desirable to maximize the number of strains/isolates sequenced so that there can be a high confidence that the broad-range signatures will indeed detect all known variations. Finishing a small number of bacterial genomes and drafting the rest will provide the best balance of sensitivity and cost for determining detection signatures. The advantages of finishing a genome are numerous,[8] but after the first representative genome for a species is completed, the benefits of completing additional strains decrease rapidly. In order to get an accurate picture of species diversity, two or three additional complete genomes may be useful, but beyond that, draft sequencing is probably sufficient. Due to the higher variability of some classes of viruses and their shorter length, finished sequence is usually obtained. Note that while draft sequencing is an excellent vehicle for mapping out SNPs and VNTRs, it is not as effective at identifying large insertions and deletions. Because a draft sequence is broken into pieces and is missing many segments, it is impossible to know whether a sequence missing from a draft represents a gap in the assembly or a genuine deletion.

It should also be stressed that obtaining sequence of close and non-pathogenic near-neighbors is a vital part of designing signatures with high specificity. While draft sequencing is generally sufficient for the purpose of isolating DNA regions specific to the pathogenic relative (see the comparative genomics discussion below), the value of finishing the nearest neighbors may be worth it. As an anecdotal example, when LLNL finished the *Yersinia pseudotuberculosis* genome, it was discovered that one *Yersinia pestis* "unique" signature was a 100% match to the non-pathogenic neighbor. Owing to a gap closure in the finishing process and the nature of the tools used to look for signature matches, the exact match of this signature had not been detected throughout the draft stage. As tools for dealing with draft genomes improve, it is probable that increased emphasis will be put on maximizing strain/isolate coverage via draft sequencing. This will make the selection of which strain(s) to finish an important choice.

STRATEGIES FOR BACTERIAL AND VIRAL SEQUENCING

The WGS method of sequencing was first applied on a large scale at TIGR and used to sequence the first bacterial organism.[9] Most labs doing large-scale bacterial sequencing now employ this approach. In some cases the genome will be fragmented into both large (10-kb) and small (1–3-kb) clones. A variety of techniques can be used to construct a *mapping scaffold* of the larger clones to aid in assembly.

Viral pathogen sequencers often employ a *parallel primer walk* strategy, where knowledge of earlier strain sequences are used to generate primer pairs to amplify overlapping chunks of the genome; these chunks may then be subjected to additional parallel primer walk sets that amplify out individual sequencing read fragments. Such strategies are clearly dependent upon the degree of strain variability, since certain mutations will render some walking primers ineffective.

ASSEMBLY

Genome sequence assembly is a complex topic about which many articles have been written; for the purposes of this chapter a brief summary will suffice. Assembly is the process of collecting together all the individual sequence reads from a WGS project and reconstructing the correct sequence of the original DNA molecule(s). For a bacterial project, this typically involves assembling tens of thousands of reads into a circular chromosome and possibly additional plasmids. Prior to the sequencing of *H. influenzae* by WGS sequencing, many scientists were skeptical that an entire genome could be correctly assembled from individual reads. Once that project, which used the TIGR Assembler,[10] demonstrated feasibility, it quickly became standard practice to assemble microbial genomes from WGS data. A few years later, WGS assembly was demonstrated to be feasible for animal genomes with the assembly of the 130 Mbp *Drosophila melanogaster* genome,[11] and soon thereafter the human genome.[12]

Although genome assemblers routinely put together bacterial-sized projects, they do not always accomplish this feat without errors. These errors are of particular concern for microbial forensics, as they may lead to mistakes in generating genetic markers. The leading genome assemblers today are designed to handle both small and large genomes, and they employ a variety of sophisticated techniques to avoid getting confused by repetitive sequences (the main cause of trouble when assembling a genome). Although the technology has

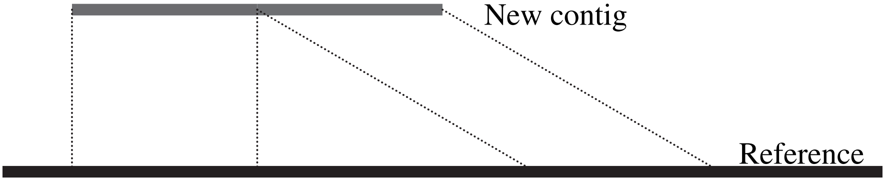


FIGURE 15.2 A newly assembled contig maps to two distinct, noncontiguous locations on a reference genome. The correctness of the new contig can be determined by PCR across the juncture. (See color insert.)

been quite successful at scaling up to larger genomes, it still has not overcome the problem of assembly errors, in part because the raw data always contain errors (but also because nearly identical repeats can be very hard to assemble correctly).

By exercising care in the interpretation of assemblies, however, these errors can nearly always be avoided. The errors generally fall into two broad categories: (1) mis-assemblies and (2) consensus errors. Let us assume for the sake of discussion that we are comparing a complete reference genome to an assembly of a related strain that has been sequenced to 8× coverage, and that the 8× assembly has produced 100 large pieces of DNA, called *contigs*. Suppose that one of these contigs appears to consist of two disjoint pieces of the reference genome that are joined together, as shown in Figure 15.2. This mis-assembly will appear to be a significant difference between the genome in question and a reference genome. There is no good method for evaluating the probability that the assembler has mis-assembled a contig; instead, the best means for testing this difference is to use PCR to test the newly-assembled contig. If PCR confirms the result, then the assembly is good and the genetic difference is real.

The second type of assembly error is a consensus error: for each position in a contig, the assembler's output includes a consensus base that represents what the majority of the underlying reads show at that position. If this does not match the genome at a given position, then it is a consensus error. This can happen for several reasons, most commonly because of basecalling errors (see the section above). Assemblies are based on a multiple alignment of the individual reads; as pointed out above, when the coverage is deep, the probability of a basecalling error is vanishingly small. In an incomplete assembly, the ends of each contig typically contain only one or two sequencing reads, and it is here that basecalling errors are most likely. Thus when comparing an assembly to a reference genome, differences that occur near the ends of contigs are likely to represent simple basecalling errors. These can be checked by additional sequencing.

The other common type of consensus error is a polymorphism, which is much more interesting in forensic or genotyping studies. If the source DNA was not extracted from a clonal organism, but instead came from two or more individuals, then the reads may contain differences that represent polymorphisms in the population. In this case, the coverage at a given position might contain a mixture of two different bases, e.g., C and T. Such polymorphic differences can be identified even without a reference genome by looking through the genome assemblies for correlated mismatches; i.e., positions where more than two reads differ from the consensus and where those reads agree with one another.

ANNOTATION

The term *annotation* is broadly used to define any special features of, or knowledge about, specific regions of sequence. Intense focus on automated gene finders has diluted the definition in many contexts to refer primarily to *gene annotation*, but it should be remembered that many non-gene features also qualify as annotation. Indeed, one could define *annotation tracks* on a pathogen genome for regions conserved across all strains/isolates, and further define those conserved regions that are apparently unique when compared to all other known sequence. Diagnostic and forensic *signature regions* derived from these conserved/unique annotation tracks could themselves form additional annotation tracks. Our discussion of annotation here will be heavily focused on aspects relevant to the design and selection of good detection and forensic diagnostics.

TOOLS FOR ANNOTATION

An outline of the approaches routinely used for prokaryotic annotation is summarized below; these are only a sample of the tools available today. The first step in most genome annotation "pipelines" is the identification of protein coding sequences. For this task, TIGR and many other centers use the *Glimmer* system,[13,14] a freely available software package that identifies genes using interpolated Markov models (IMMs), a specialized form of hidden Markov model (HMM). After an initial analysis by *Glimmer*, we use BlastP to search the predicted genes against a nonredundant database of all known proteins. We also search all intergenic regions to catch any genes that *Glimmer* might have missed (typically this is fewer than 1% of the total).

For annotation purposes and for forensic use, it is important to identify repetitive sequences in a genome. For this task one can use either *REPuter* or

MUMmer, efficient algorithms based on suffix trees.[14,15] The initial set of repeats can be further processed to group repeats into classes, which are used to guide assembly, closure, and annotation.[16]

With initial gene predictions in hand, the next major task is to predict gene function for as many genes as possible. The primary solution for this problem is to use sequence alignment, as implemented in the widely used Blast package[17] for alignment of both DNA and protein sequences. More sensitive tools for this task are based on multiple sequence alignments and gene family models[18] to assign function; in particular, HMMs built from protein families and protein domains. The PFAM[19] and TIGRFAMs[20] databases are built by creating paralogous gene families from multiple sequence alignments made with the target genome's predicted amino acid sequences. These multiple sequence alignments are then used to group similar proteins into families for verification of annotation, and to identify family members not recognized by simple pairwise alignment. A powerful feature of HMMs is that once they are built, they can be used as search targets: the efficient Viterbi algorithm as implemented in HMMer[18] allows one to search a protein or a set of proteins against an HMM and score the results. New tools that are being incorporated into the annotation process include phylogenetic analysis to aid in functional predictions[21] and "non-homology" methods for functional annotation such as phylogenetic profiling and the Rosetta stone method.[22–24] Other features of genes and genomes that can be captured or predicted by software include potential membrane-spanning domains, which are predicted by *TopPred*[25] and its more recent successors; signal peptides, predicted by *SignalP*;[26] and tRNAs, for which we use the highly sensitive and specific *tRNAScan-SE* program.[27]

The gold standard for genome annotation continues to be manual curation. Expert human curators, usually scientists with Ph.D.-level training in biology, can evaluate all the data supporting a functional assignment and can go several steps further, including reading abstracts or looking at original sources to confirm a prediction. Distilling all of this information from a wide variety of sources down to an accurate gene assignment is a complex task. Graphical display tools can provide valuable aid to these annotators; one tool for incorporating multiple types of evidence and supporting annotation is *Manatee*, an open-source manual annotation tool (freely available from manatee.sourceforge.org).

ANNOTATION FOR DIAGNOSTICS VERSUS TRADITIONAL GENOMIC ANNOTATION

In the Examples section below we describe an existing fully automated computational pipeline for whole-genome analysis to define candidate DNA signatures for bacterial and viral pathogens. Depending on the organism involved, this can yield a very few (or zero) to several tens of thousands of anonymous candidates. Despite the number, it is desirable to know the following for all candidates:

— Does it land in a gene?
— Is anything known about that gene?
— Is the gene involved in virulence?

Clearly, gene annotation is vital for the design and selection of pathogen diagnostics produced via whole-genome methods. (In contrast, traditional pathogen diagnostic design starts with a gene of known involvement in virulence.) Candidate unique diagnostic signatures that land on genes known to be involved in virulence should receive the highest priority. Experience has shown that whole-genome analysis may uncover unique signatures that land on genes with unknown function, and these too might be associated with virulence or host range selection. Some of these candidates should be taken forward for wet lab testing. It may also be prudent to randomly select some signature candidates that land on intergenic regions. These may turn out to be interesting gene regulation regions or genes that were missed by the gene caller programs. Additionally, having a few good intergenic detection signatures is a good insurance policy against genetic engineering to foil signatures targeted at obvious known virulence gene regions.

In the realm of protein detection and forensic diagnostic design, annotation plays additional roles. It can ensure that only proteins available on the organism's surface (envelope capsid or spore surface proteins) are being focused on, assuming that detection of nondisrupted organisms is desired. Knowledge of *active site* regions and protein regions that are *solvent-accessible* can guide selection of protein signature candidates.

COMPARATIVE GENOMICS

The field of *comparative genomics* is currently in a period of accelerated development triggered by the recent explosion of genome sequencing (human, model organisms, bacteria, viruses, etc.) and the subsequent need to perform comparisons between two or more genomes. Comparative genomics tools can operate on either nucleic acid or protein sequence; some tools are specific to one while others can handle both. Most people think of *alignment tools* when discussing comparative genomics, but tools that determine *common substrings* are another important category. The whole-genome computational pipelines described in the Examples below were the first large-scale application of these techniques to pathogen diagnostic development. This took advantage of many of the tools described below and has clarified the need for many additional tools. Other tools that may have similar application to the design of microbial detection and forensic diagnostics are discussed in a recent survey paper.[28]

PAIRWISE COMPARISON AND ALIGNMENT

Until very recently, most alignment algorithms were primarily designed for comparing single protein or DNA sequences, containing a single gene, either to each other or to a database of sequences. When faced with the problem of aligning long genomic sequences or entire genomes, most programs are incapable of producing accurate alignments and consume excessive space and time, although companies such as Interagon, Paracel, and Timelogic offer specialized hardware to speed up pairwise algorithms. Algorithms designed to run at higher speeds typically make a tradeoff between speed and sensitivity: faster computational time means that some alignments might be missed.

Genome-length alignment tools have usually been designed to satisfy one of several goals: some simply aim to find all similar or identical stretches of DNA between two genomes; others specifically target coding sequences (such as exons) and search for conserved exon order between two species; still others focus on intergenic and intronic regions to detect conserved regulatory signals. In the context of microbial diagnostic design, pairwise comparisons may be primarily used when a newly sequenced isolate is being compared against a very close relative. Some of the main problems associated with computing such an alignment include: (a) genome rearrangements (e.g., exon shuffling or syntenic breaks resulting from intramolecular recombinations), (b) large insertions or deletions (sequences that share several regions of local similarity separated by unrelated regions), (c) repetitive sequences (e.g., duplicated genes/operons, transposons, simple and complex repetitives), (d) tandem repeats, and (e) inherent problems of gene regulatory elements, including their small size and relative resistance to small insertions/deletions or substitutions.

MULTISEQUENCE AND MULTIGENOME ALIGNMENT

Another subject infrequently addressed for long sequences, and needing more in-depth exploration, is the issue of multiple alignments. Every alignment is based on a *colinear* arrangement of sequence similarities (e.g., the order of segments of similar sequence is preserved, although the distances between those segments may vary among the inputs). Therefore, a global alignment of whole genomes makes sense only if the species are closely related (or more precisely, if very few genome rearrangements have occurred). All global alignment programs described below rely on this assumption. Several of these programs use an *anchor*-based method, where an anchor is a subsequence shared by two or more genomes. The method is divided into three phases:

1. Computation of all potential anchors.
2. Computation of an optimal colinear sequence of nonoverlapping potential anchors: these are the anchors that form the basis of the alignment.
3. Extension of the alignment to narrow or close the gaps between the anchors.

For diverged genomic sequences, a global alignment strategy is likely to fail when it tries to align nonsyntenic and unrelated regions. In this case, one must either use local alignments or else first identify syntenic regions, which then can be individually aligned.

The era of large-scale alignment algorithms began in 1996 with the versatile alignment program called *DIALIGN*,[29,30] capable of both pairwise and multiple alignments. One of the novelties of this program was the use of gap-free whole segments for comparison, instead of using single bases or residues. The alignments are thus composed of gap-free segment pairs of equal length that would form diagonals in a dot-matrix comparison. Such segment pairs are sometimes called "fragment alignments" or "fragments." A quality score is assigned to every possible fragment based on the probability of its random occurrence, and the program tries to find a colinear collection of nonoverlapping fragments with maximum total score.

A need to compare closely related bacterial species (or even strains) motivated the creation of *MUMmer*, a pairwise anchor-based alignment program capable of detecting every difference between two microbial genomes.[14] Under the assumption that the compared sequences are closely related, this system can quickly perform high-resolution comparisons of whole-genome-length sequences, locating all the SNPs, insertions/deletions, differences in number and location of repeat elements, and tandem repeats, as well as regions repeated in only one of the two sequences. This program is also amenable to detecting the differences between two different versions of a genome sequencing project (two drafts, or a drafted genome vs. a complete one). It proceeds in the following three phases: (1) A maximal unique match (*MUM*) decomposition of the two genomes G_1 and G_2 is computed. A *MUM* is a sequence that occurs exactly once in genome G_1 and once in genome G_2, and is not contained in any longer such sequence. Using the suffix tree of $G_1\$G_2$, *MUM*s can be computed in $O(n)$ time and space, where $n = |G_1\$G_2|$ and \$ is a symbol neither occurring in G_1 nor in G_2. (2) The matches found in the *MUM* decomposition are sorted, and the longest possible set of *MUM*s that occurs in the same order in both genomes is extracted, yielding the anchors. If there are *m MUMs*, then this can be done in $O(m \log m)$ time by an algorithm which finds the longest increasing subsequence (LIS) of a sequence of weighted integers. (3) The gaps in the anchor alignment that have length less than or equal to a given limit (5,000 bp is the default in *MUMmer*) are closed with a standard dynamic programming algorithm. These gaps consist of apparent insertions/deletions (lateral transfer, transpositions), polymorphic regions, and repeated elements, which by the nature of the MUMs (U for Unique) are captured when found out of context compared with the other sequence.

MUMmer was, and continues to be, a major breakthrough toward the solution of the alignment of two sufficiently similar genomic sequences. Like most other existing methods, however, *MUMmer* cannot align more than two genomic sequences.

Recently, one of our labs released *MUMmer2*,[31] which improves on *MUMmer* by only computing the suffix tree of one sequence. This reduces the space requirement by about 50%. The other sequence is matched against the suffix tree delivering maximal matches that are unique in one sequence, but may not be unique in the other sequence. *MUMmer2* is currently unique in that it can compare an incomplete (draft) genome to either a finished genome or another incomplete genome using the companion *NUCmer* utility. Yet another program, *PROmer*, can align genomes whose proteins are similar but whose DNA is too diverse to align. This is accomplished by using a six-frame translation. A graphical interface, *DisplayMUMs*, is also freely available with the system.

Like a small number of other recently-developed programs, *MGA*[32] can produce a global multiple alignment of whole genomes of closely related species using an anchor-based method. In the first phase of *MGA*, a novel algorithm detects all *maximal multiple exact matches* (*multiMEMs*) whose length exceeds a given threshold. In short, a *multiMEM* is a sequence of length l that occurs in all genomes $G_1, \ldots G_k$ (at positions $p_1, \ldots, p_k$) and cannot simultaneously be extended to the left (left maximality) or to the right (right maximality) in every genome. The algorithm takes time $O(kn + r)$ to compute all *multiMEMs*, where r is the number of right maximal multiple exact matches. In practice, however, it suffers from the huge space requirement of the suffix tree. Consequently, *MGA*'s efficient implementation is not based on suffix trees but on enhanced suffix arrays[33] that require only 5 bytes per input character.

After computing multiMEMs and determining which ones are anchors, *MGA* closes the gaps between the anchors. First this is done by recursively applying the same method a certain number of times, thereby lowering the length threshold for the *multiMEMs*. Long gaps remain unaligned in order to cope with long insertions, deletions, etc. Short gaps are closed by a standard multiple sequence alignment program. The program *ClustalW*[34] is used for this task because it is a widely used implementation of profile-based progressive multiple alignment that has reasonable running time for short gaps.

COMMON SUBSTRING COMPARISON

Vmatch[35] is a new tool that solves a variety of large-scale string matching problems on pairs of sequence sets very efficiently. The basic concept is to preprocess a set of database sequences into an enhanced suffix array, which provides a very powerful index structure for string matching. In a recent paper by Abouelhoda et al.[33] it is shown that several string matching algorithms, originally developed for suffix trees, can be adapted to enhanced suffix arrays. The advantage of enhanced suffix arrays over suffix trees is a considerably reduced space requirement and a much faster processing in practice. *Vmatch* is accompanied by a tool, mkvtree, which computes an enhanced suffix array for a set of database sequences and stores this on file. *Vmatch* reads these files as a database and matches query sequences against the database sequences to find local similarities. To do this efficiently it utilizes a new algorithm[35] to compute maximal exact matches (MEMs), i.e. exact matches between the database sequences and the query sequences that cannot be extended further without a mismatch. Unlike the hashing methods (which first generate *k-mers* and then extend these to MEMs), the new algorithm directly computes MEMs. As a consequence, when comparing large genomes or genome sets, *Vmatch* is much faster (by a factor between 6 and 24) than previous tools utilizing the traditional hashing methods. *Vmatch* also follows the find-seed-and-extend approach, with the seeds being MEMs. These are extended using the greedy algorithm of Zhang et al.[36]

A flexible range of options and the ability to handle very large sets of sequences gives *Vmatch* the capability to efficiently compare a single draft or completed microbial genome against other microbial genomes. This is done by considering the nontarget genomes as the database sequences and the target genome as a query sequence. Utilizing an additional post processing option of *Vmatch*, one can find all substrings of the target genome over a specified threshold length that have no match in any other genome. Other options allow a degree of inexact matches to also be excluded from the output, to select those substrings in the target genome that do not even have any close match in the nontarget genome.

COMPARATIVE GENOMIC TOOLS

There are a number of tools and capabilities that are needed to be able to achieve reliable automated DNA and protein signature creation:

ALIGNING WHOLE GENOMES WITH FRAGMENT SEQUENCES

Until recently, we did not have any alignment tools that could align a set of sequence fragments against another set of fragments. The latest version of *NUCmer*, part of the *MUMmer* package, computes this problem highly efficiently. The text-based output can be displayed graphically using a tool called *MapView* released with *MUMmer3*,[37] it is now routinely used at some centers to align incomplete genome assemblies (consisting of dozens or hundreds of contigs) to other incomplete assemblies. For more distantly related species or strains, the *PROmer* program can be used instead, as was demonstrated in a comparative study of *Plasmodium yoelii* and *P. falciparum*.[38] Better graphical tools (some of which are now available in the *MUMmer3* package) to view the results of these alignments will greatly enhance their utility.

COMPARING MORE DISTANTLY RELATED GENOMES (DEALING WITH NONCOLINEARITY)

Current alignment tools are not capable of dealing with most large-scale genome rearrangements that occur among distantly related (and even some closely related) genomes. In practice, this limits our ability in some cases to analyze highly divergent families to seek DNA diagnostics that can identify all members of that family. It also hinders our ability to use multiple sequence alignments to rapidly discover common mechanisms that may have mutated significantly yet still should be recognizable as having a common origin.

FINDING COMMONALITY AMONG APPARENTLY UNRELATED OR DISTANT GENOMES (DATA MINING FOR SHARED MECHANISMS)

Our current tools achieve their tremendous speed by requiring a rather high degree of similarity between the genomes being aligned. We would like to be able (eventually) to provide hundreds or thousands of microbial genomes to a computational tool and let it discover all clusters of similar regions, regardless of the evolutionary distances involved.

MASSIVE SCALING OF ALIGNMENTS

Within this decade it will become common to have dozens or even hundreds of completed genomes for viruses. Current tools are not capable even of aligning all the HIV genomes in Genbank. (The extreme divergence, recombination, and lack of DNA repair mechanisms in HIV make it impossible for an anchor-based multiple alignment technique like *MGA* to locate common anchors, and nonanchor alignment techniques like *DIALIGN* fail on the more than 300 HIV genomes now available.) Space and time constraints need to be constantly pushed back as the cost of acquiring genomic sequence continues to plummet.

COMPARISON OF MULTIPLE DRAFT GENOMES

No tools are yet available to adequately compare multiple draft genomes, either among themselves or with completed genomes. (As noted above, *Vmatch* can find exact match substrings between such inputs, but this is only a special case of what would be desired. *MUMmer2* can compare two genomes at any stage of completion, which is a promising start for our needs.) Large genome centers currently generate sequence data at a rate roughly equivalent to one draft bacterial genome per day, and it appears likely that the scientific community will be faced with massive amounts of draft microbial sequence in the near future. A combination of the tools described above and new algorithms will be needed to accurately align all of these “draft” genomes against reference sequences (when available) or more distant species.

USE OF MASSIVE NEW COMPUTERS

Very large-scale parallel computers will soon become available for biodefense research. For example, the Lawrence Livermore National Laboratory recently announced a project with IBM that will build a computer, named Blue Gene, that will have 130,000 CPUs. To truly exploit this enormous computational power, new highly parallel alignment and analysis algorithms must be developed. It remains a challenge to divide the multiple alignment task into sub-problems which can each be solved on a single CPU, minimizing communication with other CPUs. Such algorithms would permit tackling many interesting alignment problems that could not otherwise be solved in reasonable time.

PROTEIN STRUCTURE

As detailed in the Example section below, it is possible to computationally predict which fragments of protein sequence are conserved across multiple strains/isolates and are also apparently unique when compared to all other known protein sequences. Before turning these protein signature candidates into monoclonal antibodies or ligands, it is important to know whether these fragments lie within the *active site* of the protein and thus might not be available for exploitation by a detection assay. It is also often important to know whether the unique fragment has sufficient *solvent* or *surface accessibility* for the detection assay to come into contact with the signature fragment while the protein is in a physiological conformation.

TOOLS FOR PROTEIN STRUCTURE MODELING

The structures of over 8,000 proteins have been experimentally determined and are available in the Protein Data Bank (PDB) (http://www.rcsb.org/pdb/), but these are only a tiny fraction (less than 1%) of all proteins that exist in nature. Determination of protein structures via *X-ray crystallography* or *nuclear magnetic resonance imaging* is a relatively slow and expensive process. Unlike genomic sequencing, which has seen several orders of magnitude of increase over the past several years, protein structure determination experiments have not yet experienced a quantum breakthrough.

The difficulty of increasing the rate of experimental determination of protein structures has led directly to emphasis on *computational prediction* of protein structure. The Critical Assessment of Structure Prediction (CASP) (http://predictioncenter.llnl.gov) biannual competition has seen tremendous growth in both techniques and groups involved in protein structure modeling. Both *ab initio* and *homology* methods have demonstrated ability to produce 3D models that providing additional information could be used in various circumstances. The *ab initio* methods attempt to calculate protein structure from first principles directly from the protein sequence. The basis of homology modeling methods is the conservation of structure and function among proteins exhibiting even low sequence identity, so that proteins with negligible sequence similarity can have similar three-dimensional structures. In our structure modeling work we focus on homology-based methods as still the most reliable prediction methods. Although the methods for *ab initio* prediction have developed significantly in recent years, it is still clear (as the results from CASP experiments show) that they are severely limited in terms of the size of proteins that can be modeled and the accuracy of the resulting models. In difficult protein modeling cases, the advanced homology-based methods look for regions of protein sequence homology of the target that map to regions of (possibly multiple) solved proteins in PDB, and a variety of techniques are employed to stitch these modeled regions together and analyze the range of possible models that result. In homology-based prediction methods, some or all of the following steps may be involved:

1. The set of preliminary sequence-structure alignments to known protein structures is generated using pairwise sequence alignment (Smith-Waterman,[39] FASTA3,[40] BLAST) and the multiple sequence alignment PSI-BLAST.[41]
2. A selected set of alignment-based backbone models is created.
3. The correctness of the alignment is verified by:
   a) comparison and analysis of multiple sequence alignments and structure alignments between templates,
   b) analysis of secondary structure prediction of modeled protein and comparison with secondary structure assignment from considered templates,
   c) structure comparison of all generated preliminary models.
4. Software, such as LGA[42] (Local-Global Alignment), builds regions that cannot be directly "copied" from the template structure: termini, insertions, deletions, and loops.
5. Side chains are added with the *SCWRL*[43] program. When there is more than 70% sequence identity to the closest known structure, the coordinates for corresponding atoms are inserted directly to the model.
6. The final model is constructed after an analysis and evaluation of all generated models when the completeness of the prediction and the homology level of the templates used for model building are verified (Fig. 15.3).

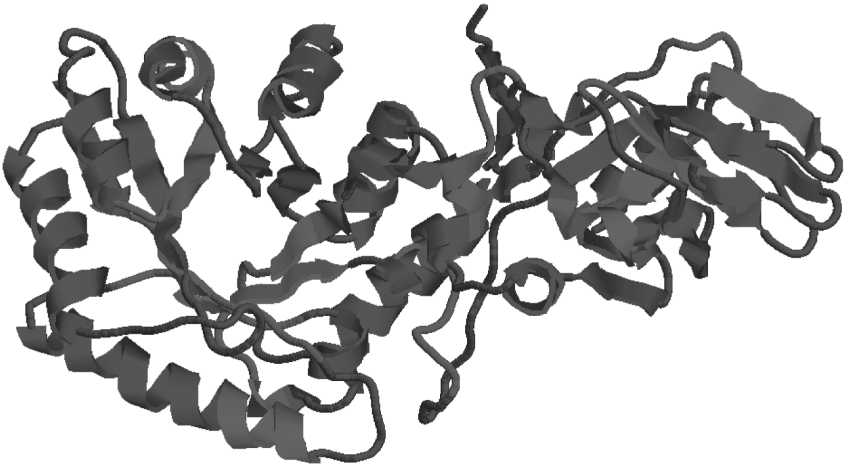


FIGURE 15.3 A structure model for a pathogen protein determined by the methods described in the text. Portions of the protein sequence are both conserved and unique, based upon currently-available sequence information. (See color insert.)

TOOLS FOR PROTEIN STRUCTURE ANALYSIS

Once we have determined protein signature candidates, we must determine which of the apparently conserved/unique protein sequence fragments are solvent-accessible (i.e., on the protein surface in normal conformation). In the absence of a 3-D model we use *PredictProtein*,[44] whereas *NACCESS*[45] is used if we do have a structure model. These programs let us filter out many candidates for monoclonal antibodies that would have a poor chance of success if taken forward for wet-lab screening.

Several protein structure analysis tools we would like to have for protein signature design either do not yet exist or are currently in early stages of development. These include:

- A tool to examine protein surface topology, looking for good sites for monoclonal antibodies that are not linear chains from the protein sequence (e.g., that lie across the top of helices or other complex formations).
- A fast tool for examining protein surface topology to locate good *pockets* for HALs that contain regions found to be conserved/unique.
- A tool using supercomputers to computationally screen large compound libraries against a set of protein pockets at a whole-proteome scale, predicting suitable HALs.

VISUALIZATION

All the techniques described above share a common need to visually describe results to the user in ways that can readily convey either a clear verdict or the reason why the answer is complex and needs further resolution with other techniques. For example, sequence and genome alignment needs to be combined with gene annotation and signature location; results of a fingerprinting technique like MLVA or SNPs needs to show the unknown sample in registration with the closest reference standards. It is worth noting that tools suitable for academic analysis and publication may not be robust enough or contain the right features for heavy-duty forensic application. Much work remains to be done to properly combine algorithms and tools into decision tools with appropriate user interfaces and visualization at each step along the way.

Similar to alignment algorithms, the direction in development of new viewing/display tools often follows the goals of the research in question. In addition to an interpretable alignment, visualization and browsing tools need to incorporate extra analyses and features such as database homologies and gene predictions from various sources. The ability to locate repetitive elements, alternate start and splice sites, protein binding sites, and other genomic features can help the biologist in these analyses.

Interactive features are other useful options to consider, such as the viewing resolution (a static graphic vs. the ability to zoom in/out) and real-time analysis capabilities (e.g., ability to search specific regions for homologies). Other problems include: (1) how or whether to represent syntenic breakpoints (e.g., genome rearrangements), (2) how to display alignments from both strands, (3) how to display multiple alignments, (4) determining whether only one sequence should be the reference for the alignment(s), and (5) how to display contigs if an unfinished genome is used as one (or more) of the entries for the alignment.

A further problem lies with the input of data for the visualization programs, since most of these were developed to work on only one specific file format, tied to the details of the research in question. Much of the work to improve the fledgling field of whole-genome comparison involves the design of new alignment algorithms and the modification or implementation of existing algorithms. These programs have often been coupled to visualization tools that try to make a seamless transition from raw data to interpretable comparisons. There remains much room for improvement in terms of long-sequence or whole-genome alignment (or multiple alignment) algorithms, in terms of formatted or processed graphical output that a user may be able to interpret and combine with other analyses.

A visualization display tool called *SynPlot*[46] was developed with the *DIALIGN* alignment algorithm in mind. *SynPlot* allows the display of multiple alignments and shows the gaps in each sequence, as well as the nature and positions of conserved regions (based on percent identity of a sliding window) for all sequences. *SynPlot* has the added functionality of being able to display the features (exons, introns, repeat elements, and CpG islands) for each sequence. As mentioned above, *MUMmer* has a display tool, and not mentioned previously was the *PIPmaker* graphical tool for displaying the output of *BlastZ* and other large-scale alignments from Webb Miller's group.[47] *MGA* has an option to output an alignment in XML format. This can be turned into HTML using the program *mga2html*. It is available, along with examples and instructions, on the *MGA* web site (http://bibiserv.techfak.uni-bielefeld.de/mga/). No visualization tool is available for *Vmatch* output, as further processing for a specific application is generally required.

OTHER FORENSIC TECHNIQUES

The proceeding discussion may make it appear that the genomics-based techniques mentioned are the best or only way to do microbial forensics. It is likely true that these techniques are the best way to define what a pathogen *is*, but that is not the whole story. Forensic identification is also concerned with such issues as *how* a pathogen was made (additives), *where* it was made (trace materials, isotopes, etc.), and *when* it was made (^{14}C dating). We will not discuss these topics here because they are discussed in other chapters and because they do not have major bioinformatics issues. Two other forensic techniques that have bioinformatics requirements will be briefly mentioned.

PROTEIN MASS SPECTROMETER ANALYSIS

The use of *mass spec* analysis is long established within classical forensics. We are all familiar with the ability of crime scene technicians to determine the make and model of a car from a paint sample. These techniques can also be applied to pathogens and to the toxins that they create in the host. A current need is to establish a reference database of *mass spec fingerprints* of common pathogens and associated toxins. This database will be useful for analyzing complex mixed samples that may contain environmental and host contaminants in addition to the pathogen and/or associated toxins. Fast and reliable software to determine the presence of any reference pathogens in such samples, and perhaps also deduce how, when, and where they came from, will be extremely useful.

IMAGE ANALYSIS

Electron microscopy is a standard tool of classical forensics applied to microbial analysis. These images can provide information about processing techniques and sophistication, additives used, preparation, delivery, and storage methods. Reference images from lab samples or actual cases need to be compared with new samples, looking for similarities that may indicate attribution. Automating these comparisons, using image processing techniques, would be a useful addition to the microbial forensics toolkit. This will not be an easy task due to the many factors (image magnification, lighting, focus, background material, etc.) that can confound scene segmentation analysis.

EXAMPLES

We have been involed in a range of activities involving microbial detection and forensics. Some of these are briefly summarized here to provide real-life examples of the concepts described above. For what should be obvious reasons we will not provide specific details of diagnostics currently in use for public health or national security.

ANTHRACIS ANALYSIS

Following the bioterrorism attacks in the U.S. in October 2001, in which multiple samples of *Bacillus anthracis* spores were sent through the mail to a variety of targets, TIGR was asked to sequence and analyze the strain of anthrax bacterium used in these attacks. The goals of the analysis were threefold: (1) to determine whether the anthrax had been genetically engineered in any way, (2) to discover any and all differences between the attack strain and known strains, and (3) to create a unique genetic signature that could be used to characterize the attack strain. Shortly after the attacks occurred, VNTR analysis showed that the strain used in the attacks was Ames, a strain originally isolated from a cow in Texas in 1981. This strain had been sent to the U.S. Army biodefense laboratory in Ft. Detrick, Maryland at that time, and subsequently distributed to multiple labs around the world engaged in anthrax research. Using all known anthrax VNTR markers, the samples were indistinguishable from other Ames isolates. At the time of the attacks, TIGR was nearly completed with the sequencing of an Ames isolate from Porton Down, England (originally from Ft. Detrick). This made it clear that one of our tasks was to determine whether or not we could discover new genetic markers that would allow us to distinguish the Porton Down strain from the attack strain.

The results of this study, published in June of 2001,[3] revealed answers to all three of the questions above. First, the attack strain was clearly not engineered: it was nearly identical to known laboratory strains. Second, we discovered 60 new genetic differences between Ames and other strains, including three insertions/deletions, eight new VNTRs, and 49 SNPs. The insertions were relatively small, and the only one large enough to contain a gene (1200 bp) was found to be occurring naturally in *Bacillus cereus*. All of these differences were validated by resequencing and by cross-checking against a panel of other Ames isolates. At least 15 of the newly discovered markers were found to vary within previously typed Ames samples, allowing us to more finely differentiate the strain using a genetic signature with these new markers.

Several of the methods described earlier in this chapter were critical in the analysis of the anthrax isolates. The two genomes were assembled using the Celera Assembler, and the resulting assemblies (the attack isolate and Porton Down isolate were at ~6× and 11× coverage, respectively) were aligned using *MUMmer2* and *NUCmer*. These assemblies contained hundreds of small and large contigs, with the Porton Down isolate containing fewer contigs due to its deeper coverage. (In addition, finishing and closure work was nearly complete at that time.) Complicating the analysis was the fact that the Porton Down isolate had been cured of its plasmids; fortunately, both the pXO1 and pXO2 plasmids had been sequenced separately, from the Sterne and Pasteur strains. So the "reference" strain was really three strains, one for the 5.2-Mbp main chromosome, and two others for the 182-Kbp and 96-Kbp plasmids.

MUMmer allowed us to very quickly identify all differences between the assemblies, which we then classified as SNPs, VNTRs, or insertions. Note that due to the time pressure on the project, assemblies were rerun several times and the analysis had to be repeated each time. The speed of the assembler and the alignment software were critical in these multiple re-analyses. We then extracted all the underlying sequence reads for every difference, and eliminated regions of 1–2× coverage as representing likely basecalling errors. For the remaining differences, we calculated the probability of error for each SNP and VNTR and reported all the high-confidence differences, along with their probabilities, in the published analysis. Subsequent sequencing of additional *B. anthracis* strains at TIGR has further confirmed the SNPs, VNTRs, and indels reported in this study: nearly all have been found in at least one additional strain or isolate. These additional findings serve to reinforce the statistical and computational methods developed for this study.

NUCLEIC ACID SIGNATURE PIPELINE

Traditional approaches to DNA signature development started with the assumption that a particular gene was vital to an organism's virulence, host range, or other factors that might be considered "unique." Suitable primers and probe were designed for the detection system of choice without effective computational screening for uniqueness. The resulting assay would then be tested with the available strain(s) and success declared if they were detected while near-neighbors were not. This approach would sometimes yield good results, but failures frequently occurred due to inadequate strain panel, near-neighbor, and environmental testing. Failure could also occur because the chosen gene was not unique as assumed but was part of some mechanism that was common to other as yet to be sequenced organisms.

LLNL embarked in early 2000 on a project to provide biosecurity for the 2002 Winter Olympics. We needed to develop robust detection assays for a range of bacterial pathogens. Our approach improved upon the traditional method by efficiently locating portions of the pathogen target genome that were not unique, and eliminated them from further consideration. In some cases, we sequenced appropriate near-neighbors to provide maximum isolation of target regions related to virulence. We used rigorous selection criteria on the remaining potentially unique portions of the genome to select candidate assays that we were confident would work against all strains of the target for which sequence was available. Finally, we worked with appropriate collaborators to ensure that our assays were screened against large strain panels and robust environmental samples. Although our approach dramatically reduced failures both in initial lab screening and subsequent field testing, the quality and quantity of target strain and near-neighbor sequence available were limiting. Thus, we view pathogen diagnostic design as a continuous process that is potentially affected by every new microbial genome sequenced.

What follows is a brief description of the system we constructed to automate the task of predicting pathogen DNA signatures using some of the tools described above. Figure 15.4 gives a simplified diagram of the KPATH system.

DETERMINATION OF TARGET PATHOGEN CONSENSUS SEQUENCE

When we began our computational DNA signature development task in August 2000, there were no algorithms available that could efficiently align multiple bacterial genomes. After experimenting with programs that did not scale to handle even tiny viruses, we began using *DIALIGN*, although it could take days or weeks to align large viral genomes. In 2002 a new multiple-genome aligner program, *MGA*, became available. This fast, anchor-based algorithm works well

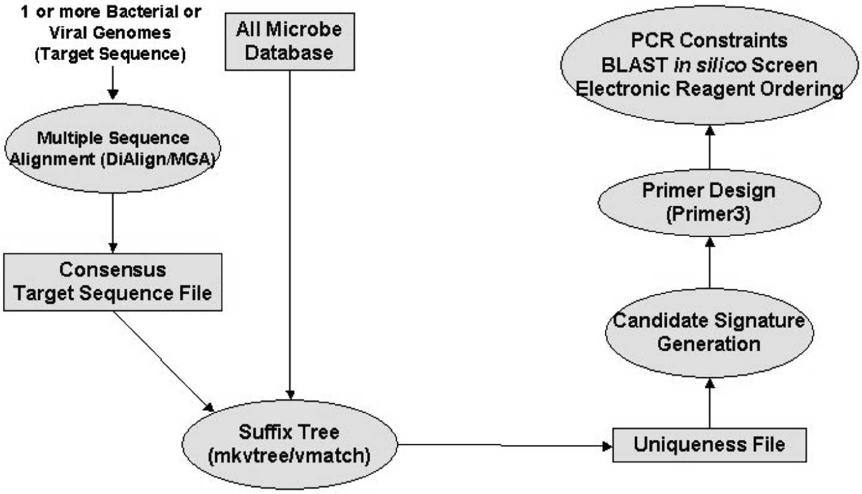

1 or more Bacterial or Viral Genomes (Target Sequence)
All Microbe Database
Multiple Sequence Alignment (DiAlign/MGA)
Consensus Target Sequence File
Suffix Tree (mkvtree/vmatch)
Uniqueness File
Candidate Signature Generation
Primer Design (Primer3)
PCR Constraints BLAST in silico Screen Electronic Reagent Ordering

FIGURE 15.4 A simplified diagram of the LLNL DNA signature pipeline.

for a collection of whole genomes similar enough to have exact-match "anchor" regions evenly distributed and present in each genome. As an example, aligning six variola (smallpox) genomes (~190 Kbp in length) took longer than a week on *DIALIGN* and required breaking each genome into three pieces. *MGA* aligned these same six genomes in less than 30 minutes. We note that all existing alignment tools assume colinearity and do not handle rearrangements or duplications well. *MGA* currently cannot align genomes too distant to have sufficient exact anchors evenly distributed, nor can it align complete genomes along with incomplete ones (i.e., draft genomes or sets of sequence fragments). We currently use *DIALIGN* and/or *MGA* to align multiple whole-genome instances of any target genome that has more than one genome available.

FAST, SCALABLE SEQUENCE COMPARISON PROGRAMS TO LOCATE UNIQUE SEQUENCE

Next, we must quickly determine which portions of a bacterial or viral genome *consensus gestalt* (Fig. 15.5) are potentially unique compared to all other microbial genomes already sequenced. We were aware that suffix trees are the most efficient data structure for comparing two strings to determine matches. Stefan Kurtz's *Vmatch* programs had just been developed and provided the most scalable implementation of suffix trees to date. He generously provided new functions that allowed us to compare a target viral or bacterial genome against a 900+ Mb library of all publicly available microbial genomes in just a few minutes. This compares to the 2–4 days required for naïve approaches (e.g., BLAST, which could involve parsing enormous output files). Note that this approach masks out portions of the target genome that are definitely *not* unique (Fig. 15.6). Further effort is needed (described below) to determine the

```
agtaatcgt.ATCATTGTACCCACTTGAGAAGTTAGTAAC.TTTTTTCTATTATAATCTT
GTATCCGTAAGATACATTACTACACATAGGAATTCCCTGAT.GAGCAATGTTTAAATACA
TCTACATTTGGAT..TGATGTAGTTGCGTATTTCTCTACAATATTAATACCATTTTTGCA
ACTATTTATTTCTAGACCTTTTG.GATTAGTAATCTCAATAATTCTACGTCAATATTATC
AGATTCTATATATTCGAATATATCAAAGTCATTGATATTTTTATAATTGGTAGAAGACAA
TAATGACACCACAACATCAGTTTTGATATTCTTATTTTT.TTGGTAACGTATACATTTAA
TGAATTTTCATTACGTTCTACCAATGATTGTGCACTGCAGGCATCAAAAGTTTTACAACT
ATCATAAAGCATACTATCCTATCC
```

FIGURE 15.5 An analysis of a multiple sequence alignment of several pathogen target genomes yields a "consensus gestalt" view. Positions that do not agree in all of the input genomes are represented by a dot. Runs of conserved positions above a threshold size are shown in capital letters; runs below that size are in lower case.

```
agtaatcgtNATCATTGTACCCACTTGAGAAGTTAGTAACNTTTTTTCTATTATAATCTT
GT.......................................N..................
..............NN............................................
........................NG..........................A.......
............................................................
..........C.............................NTTGGTAACGT.........
............................................................
..................CCT....
```

FIGURE 15.6 The consensus gestalt from Fig. 15.5 after having been compared by *Vmatch* against a database of all other sequenced microbial genomes. Positions with a dot in the consensus gestalt are mapped to "N" prior to *Vmatch* (which will not accept a dot). The dots in this "uniqueness gestalt" view indicate regions that have exact matches above a threshold size in one or more of the genomes compared with *Vmatch*. The remaining apparently-unique regions large enough to contain a PCR primer or hybridization probe are later examined for suitability as signature candidates.

suitability of the remaining portions of the target genome, and to deal with the fact that the suffix-tree approach cannot detect all regions that potentially may cross-react when hybridized even when these regions are not exact matches. We may also at times be interested in non-unique signatures (e.g., for shared virulence mechanisms). However, it is the power of being able to "mask out" all non-unique portions of a target genome by comparing it to all other microbial genomes that has enabled us to achieve a high rate of signature success compared to traditional methods. As an example, when LLNL completed the sequence of the *Yersinia pseudotuberculosis* genome, we masked out 97% of the *Y. pestis* genome, letting us focus on the 3% that was unique. The comparison library prior to the availability of this near-neighbor genome could mask out less than 33% of the *Y. pestis* genome, which illustrates the tremendous value of having a nonvirulent close-neighbor genome. Our method lets the genome itself define what is unique (and conserved, if appropriate), and presumably important.

Recently we made major improvements to our system to take advantage of multiple CPUs and allow for incremental additions to our sequence comparison database. Using a 24-CPU Sun server, we can now process all pairwise *Vmatch* comparisons for a large target bacterial genome (*B. anthracis*, for example) in less than 2 hours.

In practice, *MGA* has worked extremely well at aligning similar bacterial and (large) double-stranded DNA viral genomes. Aligning large gaps between anchors may lead to unacceptable running times. Many (short) single-stranded RNA viral genomes have such high mutation rates that *MGA* is unable to find enough anchors; in extreme cases, searching for very short anchors can exhaust memory. The requirement that all anchors must exist in all inputs is a drawback for DNA signature development instances where it would be highly desirable to align one or more finished genomes with a collection of sequence fragments (e.g., contigs from draft genomes or gene sequence fragments from Genbank).

The unique capabilities of *Vmatch* were the key to automating the task of developing pathogen DNA diagnostics. Inputs ranging from raw sequencer reads to complete 5-Mbp bacterial genomes have been compared against multigenome databases of greater than 900 Mbp using this technique. In short, a combination of efficient multiple-genome alignment and fast identification of unique substrings led to a breakthrough in the art of microbial detection diagnostics design.

PRIMER AND PROBE SELECTION AND ACCEPTANCE

Our next step is to determine regions that can meet our desired DNA detection assay format. Currently we use TaqMan® PCR, which consists of a forward and reverse PCR probe (for amplification) and an internal oligo probe with a fluorescent tag for detection. Many bioinformatics tools are available to determine the selection of hybridization primers and internal oligo probes. After much experimentation, we rejected all tools that lacked the ability to run in batch mode via a command-line interface, and settled on the *Primer3* program from MIT.[48] We tuned it to the precise needs of our chosen platform, the Cepheid Thermal Cycler (http://www.cepheid.com). These parameters include the desired amplicon length range, optimal primer and probe sizes and spacing, annealing temperature ranges, and so on.

We developed custom software to automatically extract all potential signature candidate regions from the "masked out" target genome *uniqueness gestalt* (Fig. 15.5) produced from the *Vmatch* processing described above. Each candidate region is input to *Primer3* and the output examined for meeting our criteria. Successful candidates are then subjected to an even more rigid set of local PCR primer design criteria that help us to screen out primers that might fail or internal oligo probes that would not meet strict TaqMan requirements. Candidate signatures that survive this electronic gauntlet have a high probability of detecting the target genome successfully and not cross-reacting with any already sequenced genome. Like the prior step, signature candidate selection is now done in parallel. A large target bacterial genome without sequenced, non-pathogenic near-neighbors can generate 30,000 or more signature candidates. Thus, our *in silico* process of winnowing down to the most promising signature candidates saves substantial money and time at the bench roughly proportional to the available sequence for target and near-neighbors.

We have used these techniques described above to develop nucleic acid detection signatures for several dozen bacterial and viral pathogens in the last two years. Many of these core assays have undergone rigid CDC validation assays and have entered the public health network through the CDC's Laboratory Response Network (LRN) (http://www.bt.cdc.gov/lrn/) to be used for regular pathogen detection monitoring in the BioWatch program (http://www.llnl.gov/str/October03/Imbro.html).

PROTEIN SIGNATURE PIPELINE

With straightforward extensions, the techniques above for nucleic acid detection diagnostics can be applied to protein sequences. When multiple proteomes are available from different strains or isolates, we perform a multiple sequence alignment on each protein to determine a protein consensus, using *DIALIGN*. These are compared using *Vmatch* against the GenBank nonredundant protein database, *nr*. This output is analyzed for apparently unique peptides of a given length (six or more amino acids, normally, in our application). Longer unique peptides are favored if, for example, all their substrings of length 6 are also unique.

We focus on proteins known, through annotation, to be presented on the outer surface of the pathogen in the form in which we expect to encounter it (e.g., spore in the case of anthrax). The protein structure prediction method described above is used to attempt to produce a 3-D model, since few pathogen virulence proteins have had structures determined experimentally. We must then determine which of the apparently conserved/unique protein sequence fragments are on the protein surface in normal conformation. We use *NACCESS* if we have been able to determine a structure model. We eliminate from further consideration all candidate monoclonal antibody targets that are not sufficiently exposed. Our protein signature development capability now far exceeds the speed at which monoclonal antibodies can be created for testing in the wet lab.

Figure 15.7 shows the result of this process for one pathogen protein. Regions shown in green indicate conserved and unique peptides that are accessible on the surface of the protein, and thus may provide good antibody detection targets. Potential applications of this process to determine vaccine or therapeutic targets, in addition to detection diagnostics, should be apparent. Additional work in progress at LLNL, requiring the use of very high-end parallel computers, is attempting to mine this information to automatically determine good HAL targets and, ultimately, reagents. A target goal would be to sequence a new pathogen in a day, annotate in a day, develop nucleic acid diag-

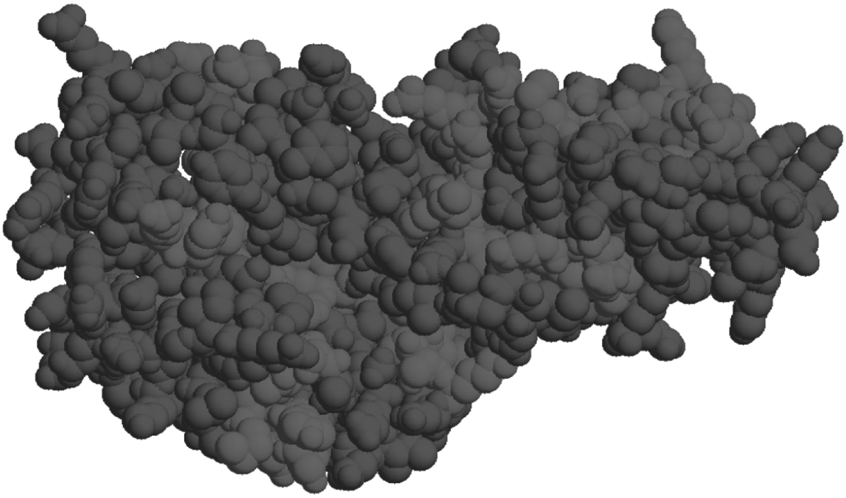


FIGURE 15.7 A 3-D model of a pathogen protein, highlighting conserved and unique protein sequence peptides that are accessible on the protein surface. These indicate potential locations for protein detection signatures. This example shows how the basic structure model shown in Fig. 15.3 can be further developed. (See color insert.)

nostics in a day, model all proteins in a day, and develop protein diagnostics before the end of the week.

ACKNOWLEDGMENTS

We thank Burkhard Morgenstern, Stefan Kurtz, and Enno Ohlenbusch for providing technical descriptions of *DIALIGN*, *MGA*, and *Vmatch*. Patrick Chain, Carol Zhou, Adam Zemla, and Elizabeth Vitalis provided considerable assistance improving this chapter. S.L.S. was supported in part by NIH grants R01-LM06845-04 and NO1-AI-15447. T.S. was supported by the U.S. Department of Energy at the University of California, LLNL under Contract No. W-7405-Eng-48.

REFERENCES

1. Keim P, Price LB, Klevytska AM, et al. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus* anthracis. *J Bacteriol* 2000;182(10):2928–2936.
2. Klevytska AM, Price LB, Schupp JM, et al. Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. *J Clin Microbiol* 2001;39(9):3179–3185.
3. Read TD, Salzberg SL, Pop M, et al. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus* anthracis. *Science* 2002;296(5575):2028–2033.

3a. Pop M, Phillippy A, Delcher AL, Salzberg SL. Comparative genome assembly. *Briefings in Bioinformatics* 2004;5(3):237–248.

4. Lander ES, Waterman MS. Genomic mapping by fingerprinting random clones: A mathematical analysis. *Genomics* 1988;2:231–239.
5. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998;8(3):175–185.
6. Altshuler D, Pollara VJ, Cowles CR, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 2000;407(6803):513–516.
7. Gajer P, Schatz M, Salzberg SL. Automated correction of genome sequence errors. *Nucleic Acids Res* 2004;32(2):562–569.
8. Fraser CM, Eisen JA, Nelson KE, et al. The value of complete microbial genome sequencing (you get what you pay for). *J Bacteriol* 2002;184(23):6403–6405; discussion 6405.
9. Fleischmann RD, Adams MD, White O, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995;269(5223):496–512.
10. Sutton G, White O, Adams M, Kerlavage AR. TIGR Assembler: A new tool for assembling large shotgun sequencing projects. *Genome Science and Technology* 1995;1:9–19.
11. Myers EW, Sutton GG, Delcher AL, et al. A whole-genome assembly of Drosophila. *Science* 2000;287(5461):2196–2204.
12. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science* 2001;291(5507):1304–1351.
13. Salzberg SL, Delcher AL, Kasif S, White O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* 1998;26(2):544–548.
14. Delcher AL, Kasif S, Fleischmann RD, et al. Alignment of whole genomes. *Nucleic Acids Res* 1999;27(11):2369–2376.
15. Kurtz S. Reducing the space requirement of suffix trees. *Software Practice and Experience* 1999;29(13):1149–1171.
16. Volfovsky N, Haas BJ, Salzberg SL. A clustering method for repeat analysis in DNA sequences. *Genome Biol* 2001; 2(8):RESEARCH0027.
17. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol* 1990; 215:403–410.
18. Eddy SR. Profile hidden Markov models. *Bioinformatics* 1998;14(9):755–763.
19. Bateman A, Birney E, Durbin R, et al. The Pfam protein families database. *Nucleic Acids Res* 2000;28(1):263–266.
20. Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families. *Nucleic Acids Res* 2003;31(1):371–373.
21. Eisen JA. Phylogenomics: Improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res* 1998;8(3):163–167.
22. Marcotte EM, Pellegrini M, Ng HL, et al. Detecting protein function and protein-protein interactions from genome sequences. *Science* 1999;285:751–753.
23. Suhre K, Claveire JM. Fusion DB: a database for in-depth analysis of prokaryotic gene fusion events. *Nucleic Acids Res* 2004;32(database issue):D273–D276.
24. Pellegrini M, Marcotte EM, Thompson MJ, et al. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci USA* 1999;96:4285–4288.
25. Claros MG, von Heijne G. TopPred II: An improved software for membrane protein structure predictions. *Comput Appl Biosci* 1994;10(6):685–686.
26. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 1997;10(1):1–6.
27. Lowe TM, Eddy SR. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997;25(5):955–964.

28. Chain P, Kurtz S, Ohlebusch E, Slezak T. *An applications-focused review of comparative genomics tools: Capabilities, limitations, and future challenges. Briefings in Bioinformatics* 2003;4(2):133–149.
29. Morgenstern B, Dress A, Werner T. Multiple DNA and protein sequence alignment based on segment-to-segment comparison. *Proc Natl Acad Sci USA* 1996;93:12098–12103.
30. Morgenstern B, Frech K, Dress A, Werner T. DIALIGN: Finding local similarities by multiple sequence alignment. *Bioinformatics* 1998;14(3):290–294.
31. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* 2002;30(11):2478–2483.
32. Hohl M, Kurtz S, Ohlebusch E. Efficient multiple genome alignment. *Bioinformatics* 2002; 18(Suppl. 1):S312–S320.
33. Abouelhoda MI, Kurtz S, Ohlebusch E. The Enhanced Suffix Array and Its Applications to Genome Analysis. Proceedings of the Second Workshop on Algorithms in Bioinformatics, Vol. 2452: Springer-Verlag, 2002. pp. 449–463.
34. Thompson JD, Higgins DG, Gibson TJ. CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 1994;22(22):4673–4680.
35. Kurtz S. A Time and space efficient algorithm for the substring matching problem. *Zentrum für Bioinformatik, Universität* Hamburg, 2003.
36. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. *J Comp Bio* 2000;7(1/2):203–214.
37. Kurtz S, Phillippy A, Delcher AL, et al. Versatile and open software for comparing large genomes. *Genome Biol* 2004;5(2):R12.
38. Carlton JM, Angiuoli SV, Suh BB, et al. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* 2002;419(6906):512–519.
39. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol* 1981;147(1):195–197.
40. Pearson WR. Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol Biol* 2000;132:185–219.
41. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 1997;25(17):3389–3402.
42. Zemla A. LGA: A Method for finding 3-D similarities in protein structures. *Nucleic Acids Res* 2003:31(13):3370–3374.
43. Bower MJ, Cohen FE, Dunbrack RL, Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modeling tool. *J Mol Biol* 1997; 267(5):1268–1282.
44. Rost B, Sander C. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 1994;19(1):55–72.
45. Hubbard SJ, Eisenmenger F, Thornton JM. Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci* 1994;3(5):757–768.
46. Gottgens B, Gilbert JG, Barton LM, et al. Long-range comparison of human and mouse SCL loci: Localized regions of sensitivity to restriction endonucleases correspond precisely with peaks of conserved noncoding sequences. *Genome Res* 2001;11(1):87–97.
47. Schwartz S, Zhang Z, Frazer KA, et al. PipMaker—A web server for aligning two genomic DNA sequences. *Genome Res* 2000;10(4):577–586.
48. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000;132:365–386.

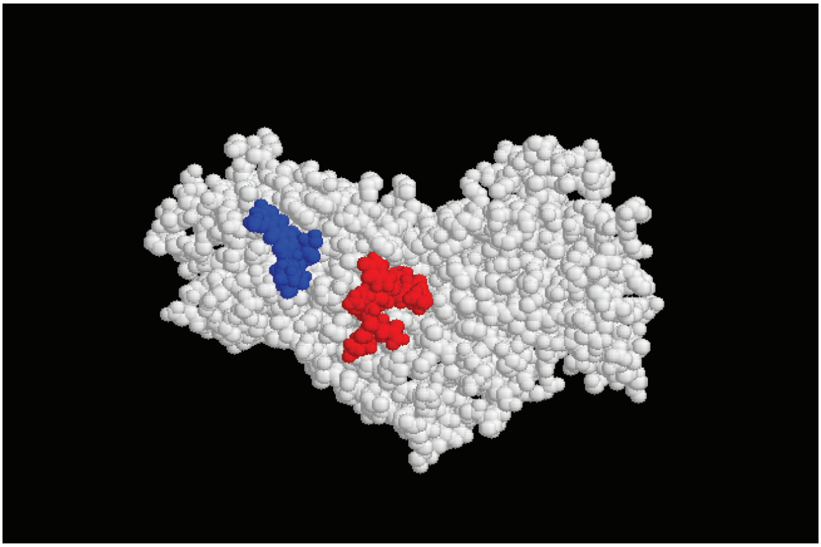


FIGURE 15.1 A representation of a bi-dentate ligand that recognizes the botulinum toxin. Figure courtesy of Rod Balhorn, LLNL.

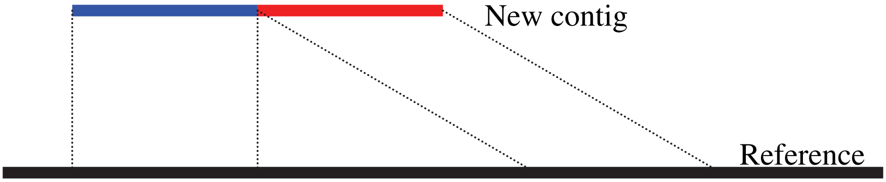


FIGURE 15.2 A newly assembled contig maps to two distinct, noncontiguous locations on a reference genome. The correctness of the new contig can be determined by PCR across the juncture.

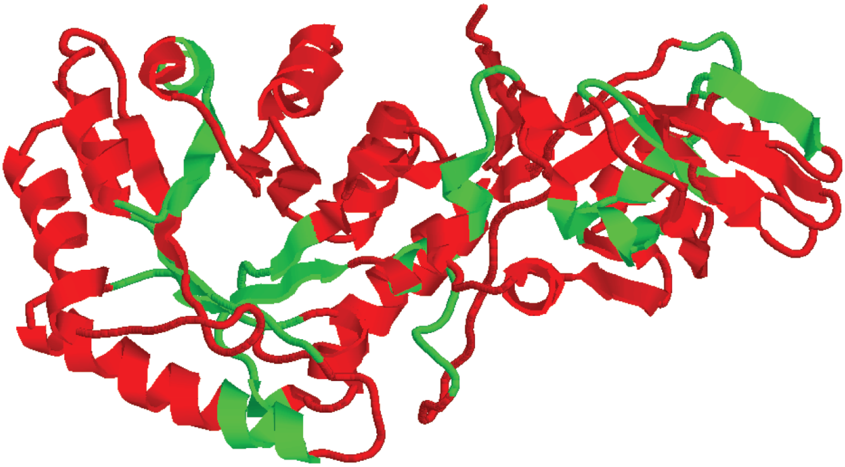


FIGURE 15.3 A structure model for a pathogen protein determined by the methods described in the text. Regions in green indicate portions of the protein sequence that are both conserved and unique, based upon currently-available sequence information.

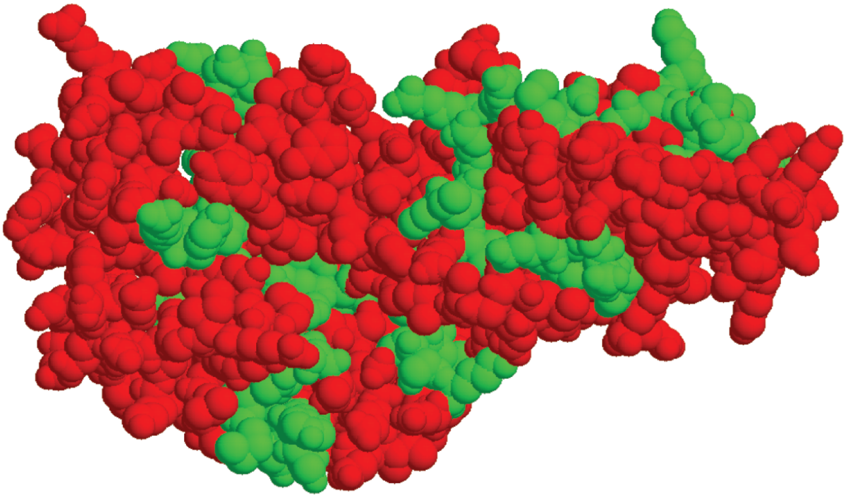


FIGURE 15.7 A 3-D model of a pathogen protein, highlighting conserved and unique protein sequence peptides that are accessible on the protein surface. These indicate potential locations for protein detection signatures. This example shows how the basic structure model shown in Fig. 15.3 can be further developed.

CHAPTER 16

Population Genetics of Bacteria in a Forensic Context

RICHARD E. LENSKI

Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan

PAUL KEIM

Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona

INTRODUCTION

One of the central goals of microbial forensics is to identify the source of a microorganism that has been used for terrorist or other illicit purposes. To achieve this goal, it is necessary to understand the extent of microbial diversity, the genetic processes that generate diversity, other evolutionary processes that shape the patterns of diversity, and the resulting genealogical relationships among diverse strains. The field of population genetics is concerned with precisely these issues and is therefore central to microbial forensics. Population genetics can be viewed as a subfield of evolutionary biology that focuses on genetic diversity within species, as opposed to differences between species and higher taxonomic groups, although this distinction is not always precise, especially in microbes.

An understanding of the population genetics of the human species already plays a major role in forensics. For example, by comparing the DNA "fingerprint" in a sample of tissue (e.g., blood) from a crime scene with the DNA from a suspect in that crime, one can say with a high degree of statistical certainty whether the suspect was the source of the forensic sample. Within an appropriate evidentiary context, a match in such a comparison provides powerful evidence of the suspect's involvement, whereas the absence of a match can exonerate an accused individual.

The power of DNA fingerprinting for forensics was not a forgone conclusion: instead there was intense debate for several years.[1] The eventual consensus that supported the utility of human DNA in forensics was reached after

Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

systematic collection of data on the extent of diversity at the genetic loci used in testing, as well as detailed analyses of how that genetic diversity was distributed among different populations. This research has led to a greater basic understanding of the population genetics of the human species, as well as to improved forensic methods.

Given the successful forensic application of population genetic analyses of human data, it is not surprising that similar approaches are being pursued to trace the source of microorganisms (including bacteria, viruses, and fungi) whose genetic material is present in forensic samples. In fact, such work also represents a direct extension to forensics of the approaches that are widely used by molecular epidemiologists to track the source of outbreaks of many pathogens.

There are important differences in population genetics between humans and microorganisms. The aims of this chapter are to explain these differences, illustrate ways in which useful forensic inferences might be drawn from microbial DNA sequences, and suggest avenues for future research. Because microorganisms are themselves extraordinarily diverse, we will focus on bacteria and, as appropriate, use data on *Bacillus anthracis* to illustrate certain calculations. To start, we will compare and contrast inferences that can be drawn from DNA-based forensic evidence derived from humans versus bacteria.

DNA FORENSICS OF HUMANS AND BACTERIA

All humans share the vast majority of their DNA sequences.[2–4] Nonetheless, with genomes containing billions of nucleotide base-pairs (bp), and with sexual reproduction scrambling the variants every generation, no two humans are identical throughout their genomes, with the rare exception of identical twins (equivalent to exact clones). With enough high-quality sequence data, therefore, the source of a human forensic sample can be attributed to a particular individual with certainty. Obtaining enough data to make a strong probabilistic argument is made easier because some regions of the human genome are hypervariable,[5] allowing analyses to focus on those regions rather than requiring whole-genome sequences. This individuality of the genetic signature gives rise to the metaphor of DNA fingerprinting.

In extending the use of DNA evidence from human samples to bacterial DNA forensics, one possible line of reasoning might be as follows. Most bacterial species harbor tremendous sequence diversity. With bacteria having much smaller genomes than humans, it also becomes feasible to obtain full genome sequences for forensic samples. Therefore, according to this view, it should be easier and more certain to trace back from a bacterial sample to its source than to do so in the case of human DNA forensics. Unfortunately, this reasoning is, at least in some cases, false owing to the greater potential for exact clones (equivalent to identical twins) in bacteria than in humans.

Bacteria reproduce asexually, hence the existing genetic diversity within a population or species is not scrambled every generation. Moreover, mutation rates in bacteria are generally not high enough to ensure that new mutations occur in every cell generation.[6,7] Thus, exact clones are much more prevalent in the bacterial realm than in the human population.

Furthermore, microbiological research laboratories often preserve exact or nearly exact clones, and may distribute them to other laboratories. Hence, the possibility of exact clones is especially important when considering pathogens that may derive, accidentally or deliberately, from a laboratory. By contrast, when tracing the source of a natural outbreak, exact clones are likely to be less problematic because the number of generations and accumulated mutations are usually much greater, the relevant sources are not deliberately stored somewhere as clones, and perfect attribution is often less critical.

CASE STUDY OF *BACILLUS ANTHRACIS*

In this section, we will present some quantitative considerations related to microbial forensics. We will use published research on *Bacillus anthracis* to illustrate several important issues. We do so because the anthrax terrorist attacks of October 2001 have received considerable attention and, thus, offer concrete data for discussion. However, we will gloss over certain complications of this particular example in order to emphasize more general issues that would probably apply to similar cases in the future.

Following the anthrax attacks, Read, et al.[8] sequenced and compared the complete genomes of two *B. anthracis* isolates. One isolate was a forensic sample taken from a victim in Florida, while the other was from the government research laboratory at Porton Down in the United Kingdom. The Porton Down isolate had been previously "cured" of the two extrachromosomal plasmids that encode virulence factors, but otherwise it was presumed to be representative of the Ames strain that had been widely used in anthrax research. These two isolates, as well as other potential sources of the Florida attack strain, are shown in Figure 16.1. These other sources include U.S. governmental laboratories and field isolates.

Read, et al. reported that "Only four differences were discovered between the main chromosomes of the Florida and Porton isolates . . . two of these are SNPs and two are short indels." SNPs refer to single-nucleotide polymorphisms, which in this context are typical point mutations that distinguish the two sequences. Indels are insertions or deletion mutations, which often occur in specific hypermutable regions. We will focus initially on the two point

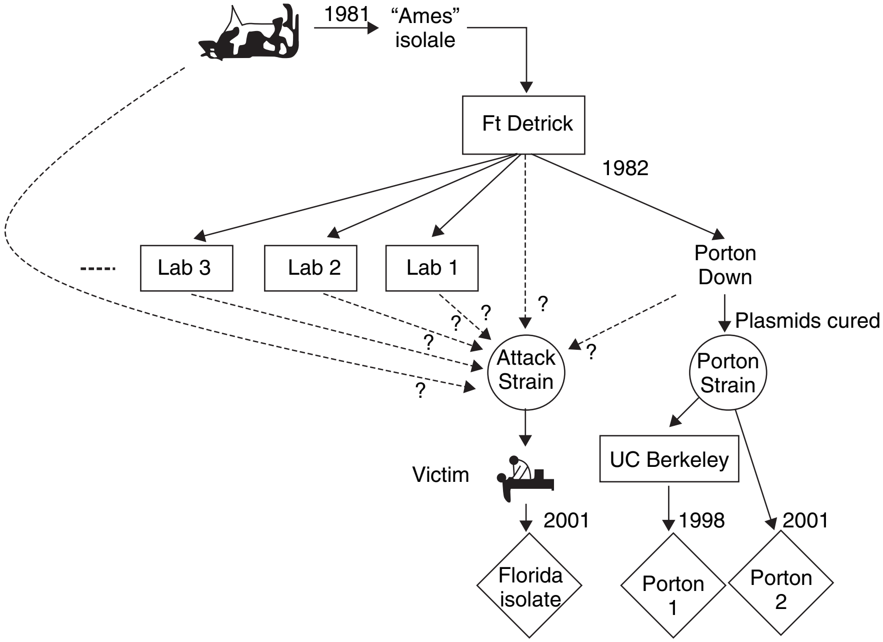

1981
"Ames" isolale
Ft Detrick
1982
Lab 3
Lab 2
Lab 1
Porton Down
Plasmids cured
?
Attack Strain
Porton Strain
Victim
UC Berkeley
2001
1998
2001
Florida isolate
Porton 1
Porton 2

FIGURE 16.1 Possible derivations of the *Bacillus anthracis* isolated from the Florida anthrax victim in relation to several potential sources of the Ames strain. Chromosomes of the Florida isolate and the Porton Down laboratory strain have been sequenced. The Ames strain was originally isolated from a dead cow in Texas in 1981, then stored at Fort Detrick, and distributed to several other research laboratories (solid arrows), including Porton Down. The Porton Down strain was cured of its two virulence plasmids. (Reprinted with permission from ref. 8. Copyright 2002, American Association for the Advancement of Science.)

mutations that distinguish the Florida and Porton Down isolates. A complication is that the genome sequence of the Porton Down isolate was, in fact, based on DNA preparations from two substrains of the Porton Down strain, and these substrains were shown to have several mutational differences in their sequences. In our analysis, we will use only those differences that distinguish the Florida isolate from both Porton Down samples.

Given that there are two point mutations that distinguish the Florida and Porton Down isolates, we can ask several questions. Is that a surprisingly little difference, or is it a lot? What might these data tell us about how long ago the forensic isolate shared a common ancestor with the Ames laboratory strain? What might the data say about the relative likelihood of the forensic isolate coming from one source versus others?

To begin to answer these questions, we first need to understand the genomic mutation rate and the evidence concerning this rate in bacteria. The genomic mutation rate is simply the expected number of mutations per generation across the entire genome. To illustrate, consider *Escherichia coli*, which is the best studied bacterium. Our calculations assume functional DNA repair and ignore hypermutable sites, which constitute a small proportion of the genome. *E. coli* has a genome size of about 5×10^6 bp and a point mutation rate of about 5×10^{-10} per bp per generation[6] (see ref. 7 for a somewhat lower estimate). The product of these two quantities is the total genomic mutation rate, which in this case is approximately 2.5×10^{-3} point mutations per generation. The inverse of the genomic mutation rate is the expected number of generations until the first mutation occurs in a cell lineage (not in a population: see below), in this case about 400 generations. As it turns out, *B. anthracis* has a similar genome size,[9] and its mutation rate measured for a particular gene also is similar to *E. coli*.[10]

At first glance, the expected time of 400 generations until the first mutation may seem an inappropriately long period, because it ignores the fact that a bacterial population may contain millions of cells, such that many mutations can occur every generation. However, comparisons between genomic sequences are based on single representatives of each sample, not on entire populations. Without belaboring this point, the expected time that is relevant to our analysis will generally be somewhat longer, not shorter, than we estimated above.

In point of fact, we are most interested in a quantity called the genomic substitution rate. A substitution is any mutation that spreads throughout a population of interest. The details can get complicated quickly but, fortunately, there are some mathematical shortcuts. In the present context, we can treat the number of substitutions as the number of mutations that distinguish the two individuals whose genomes are under comparison. Neutral mutations have no effect on fitness; synonymous point mutations are often used as a proxy for neutral mutations. A robust result from theoretical population genetics is that the expected substitution rate of the class of neutral mutations is equal to their corresponding mutation rate.[11] Deleterious mutations—those which reduce a cell's survival or growth rate—have a substitution rate lower than the corresponding mutation rate, while beneficial mutations have a substitution rate above their mutation rate. Because many more mutations are deleterious than are beneficial, the genomic substitution rate is, in general, somewhat lower than the genomic mutation rate. Hence, as noted above, the expected time to the first substitution of a point mutation will be longer than 400 generations.

So what might we begin to conclude from the data? Given the two point mutations that distinguish the Florida and Porton Down isolates, the inferred time since their common ancestor is on the order of 800 cell generations (i.e., twice the expected time to the first mutation). Although not a huge number, it is to us surprisingly large in the context of standard lab practice, where one would expect working subcultures to be repeatedly restarted from a master culture (stored in a non-growing state, as spores or frozen vegetative cells), rather than by sustained propagation of subcultures. The inferred 800 generations would correspond to about 30 rounds of plating for single colonies (each colony representing some 25 cell divisions), and even more rounds if cells were propagated by serial dilution and transfer. Translating generations into chronological time would further depend on knowing how long cells might have been stored in a non-mutating state, for example as spores or in a freezer.

The inferred time since the common ancestor could be reduced, perhaps dramatically, if the Ames strain were defective in DNA repair (see refs. 7, 12 for the effects of loss of repair in *E. coli*) or if there was a history of mutagenesis. In terms of DNA repair, it appears that the Ames strain retains these functions given that mutation-rate estimates at specific loci are in line with estimates for *E. coli* that have normal DNA repair functions.[10] However, with respect to growth under mutagenic conditions, it could be relevant that curing the two virulence plasmids from the Porton Down strain involved treating cells with high temperature and an antibiotic.[9] In fact, depending on the timing of these treatments relative to the derivation of the two substrains of the Porton Down strain, these conditions may even explain the differences between the substrains as well as between the Florida isolate and the Porton Down strain.

Statistical uncertainties are also important with respect to inferring the time since two strains diverged from a common ancestor. Even if we accept the substitution rate as known, there remains the intrinsic error arising from the stochastic (random) occurrence of mutations. Using the Poisson distribution to reflect this intrinsic error, the probability of observing two or more mutations in anything fewer than 142 generations is below 5%. At the other end of the distribution, two strains are also unlikely to have substituted as few as two mutations in 1900 generations or longer ($p < 5\%$). Although these bounds are already large, they are the best that one can do given only two point-mutation differences, because they assume that all else is known precisely. The bounds become even larger with uncertainty in, for example, the mutation rate. Despite these wide bounds, one might still exclude certain scenarios. For the sake of illustration (leaving aside the facts that the Porton Down strain lacks the virulence plasmids, and was subjected to mutagenic treatments), these two point mutations make it unlikely that the forensic isolate came directly from the Porton Down strain or even that it was derived from that strain via fewer than several rounds of plating or subculturing.

Thus, after considering what is known about rates of mutation in bacteria and placing this information in an evolutionary context, the two point mutations separating the forensic isolate from Florida and the Ames laboratory strain appear to be more than might have been reasonably expected, not less, provided that the Porton Down strain is indeed representative of the Ames strain more generally. We will return to this proviso a bit later.

It should also be clear from what has been said that "either-or" inferences based on match versus no-match between a possible source and a forensic sample are less conclusive for bacteria than for humans, owing to the much greater possibility of exact clones (identical twins) in the microbial case and especially in the context of a deliberate attack using a strain taken from the laboratory.

Let us now shift gears, and consider these data from the perspective of establishing the most probable "line of descent" of a forensic isolate in relation to multiple potential sources. In this context, even one or a few distinguishing genetic substitutions could—in principle—provide compelling evidence to support or exclude certain scenarios. In the paragraphs that follow, we examine two such scenarios to illustrate how the data could be profitably analyzed. The first scenario is hypothetical and invokes imaginary data in order to make certain points clear. The second scenario accords with the relevant published data.

FIRST SCENARIO

Recall that the forensic isolate from Florida and the Porton Down version of the Ames strain differ by two point mutations. We can designate the genotype of the Porton Down isolate as AB and that of the Florida isolate as A′B′ to reflect these two mutations. Now suppose that isolates of the Ames strain from the four laboratories in Figure 16.1 are also characterized with respect to these mutations, along with an isolate from a hypothetical rogue laboratory discovered in an investigation. Suppose that the Ames isolates from Fort Detrick, Lab 1 and Lab 2 had the same AB genotype as the Porton Down strain, while the isolate from Lab 3 yielded the A′B genotype and the isolate from the rogue laboratory had the same A′B′ as the forensic isolate from Florida.

Under this hypothetical scenario, the identity of the isolate from the rogue laboratory with the forensic sample, coupled with the differences between these isolates and those from all other laboratories, would suggest that the likely source of the forensic isolate was the rogue laboratory. Moreover, the fact that the isolate from Lab 3 alone shared one of the two distinguishing point mutations would further suggest that the strain taken from the rogue lab was derived from Lab 3. Of course, DNA sequences of anthrax isolates would presumably not be the only line of evidence presented in a criminal case against the hypothetical operator of the rogue laboratory; the case would be further strengthened, for example, if the operator had previously worked in Lab 3.

At first glance, the finding that these two point mutations are shared by the forensic and rogue-lab isolates may not seem like much, given that the *B. anthracis* genome has about five million bp. However, when viewed against a background of genomic uniformity, even one or a few shared mutations provide compelling quantitative support for an association. To illustrate, we will make some simplifying assumptions and rough calculations. No doubt such assumptions could be relaxed, and the calculations refined, in any actual case.

If all mutational substitutions were equally likely, then the probability that some putative source would have independently substituted the exact same m mutations (and no others) as in a forensic sample is calculated as $p = 1/(3n)^m$, where n is the genome size in bp and the factor of 3 represents three alternative base-pairings at each genomic site. With a genome of $n = 5 \times 10^6$ bp and $m = 2$ mutations, this probability of chance convergence is $<10^{-14}$.

However, the refined probability estimate would be greater owing to variation among genome positions in substitution probabilities. Certain positions are more mutable than others,[13] and therefore coincident mutations become more probable than the calculation above suggests. Also, selection can sometimes lead to convergent substitutions even when the underlying mutation rates are the same.[14] To illustrate, imagine that 1/1000th of the genome (5×10^3 bp) is highly mutable, with all of these sites equally mutable and with unrestricted nucleotide substitution. In that case, the probability of chance convergence with $m = 2$ is still $<10^{-8}$. If substitutions at each of the highly mutable sites are restricted to one particular nucleotide, then the probability of convergence is higher (as $p = 1/n^m$), but still $<10^{-7}$. In any case, it is important to emphasize that a few, or even one, mutational matches could be quantitatively compelling, especially if the sequence data are supported by other evidence (such as the hypothetical connection between the rogue laboratory operator and a related source strain).

SECOND SCENARIO

In fact, when the mutations that distinguish the forensic isolate and the Porton Down strain were checked among isolates from the four other laboratories in Figure 16.1, it was found that these other isolates all shared the same genotype as the forensic sample.[8] Thus, it seems the Porton Down strain accumulated these distinguishing mutations, probably during the mutagenic treatments used to eliminate the virulence plasmids.[9]

More importantly, the forensic isolate appears to be identical in its genomic sequence to the Ames strain that was shared among several different laboratories. It is possible that complete genome sequences of isolates from one or more of these laboratories might exclude one of them as the proximate source. However, examination of hypermutable sites [variable-number tandem repeats (VNTRs)] in the genomes of these isolates does not reveal any exclusionary differences.[8] Thus, the genomic data per se lack the power to discriminate between these particular laboratories as potential sources of the Florida forensic isolate. Although it is not a focus of this chapter, we should mention that other types of evidence, such as trace elements reflecting bacterial growth conditions, might allow source discrimination even in the absence of distinguishing mutations. Of course, evidence concerning access to *Bacillus anthracis* by a suspect individual or group would also be quite relevant.

In Figure 16.1, one can also see that the Ames strain found in all of the laboratories was itself derived from a dead animal in 1981. This source also raises the possibility that the Florida isolate was not obtained from any of the research laboratories but, instead, was another isolate sampled from nature by the perpetrator. (In this case, the possibility of a natural infection was excluded by the circumstances of the deliberate attacks using the postal system.) The question then becomes: What is the probability that a fresh isolate taken from the wild would be an exact clone of the isolate sampled twenty years earlier? In bacteria that cannot form spores, one might estimate the number of elapsed generations and the expected number of substitutions that would have accumulated, and use these to calculate the likelihood of zero mutational substitutions. In fact, this likelihood would be the upper bound because it would assume that the same evolving lineage was chosen in both instances and thus ignores the extent of genetic diversity present at any moment. However, *B. anthracis* forms spores that can persist for a long time, making it difficult to estimate the number of elapsed generations and hence expected substitutions.

An alternative approach would be to estimate empirically the fraction of isolates from nature that are exact clones of the Ames strain. At first glance, this might seem a daunting task that would require genome sequences from hundreds of isolates. In fact, a shortcut exists by first scoring the hypermutable loci. For example, one could sample isolates from various geographic regions, including that where the Ames strain was sampled. All of these could be scored for the hypermutable genes, and any differences would exclude identity. Let us imagine, for the purpose of illustration, that none of 500 isolates from other geographical regions matched the Ames strain at all of these genes. In that case, the probability of a match from outside the region would be estimated as below 0.2%. Now imagine also that five isolates out of 100 sampled from the vicinity of the Ames strain matched that strain at all of the hypermutable loci. The probability of a match from within that region might first be estimated as 5%, but it must be emphasized that this estimate is likely to be an overestimate because it uses only a subset of the genome (i.e., the hypermutable loci) to establish differences. One might then sequence whole genomes of the five isolates that matched the Ames strain at all of the hypermutable genes to see whether any of them was truly an exact clone. Alternatively, various other genetic approaches less costly than whole-genome sequencing might be used to screen this subset of clones. Among our recommendations, we suggest efforts to develop methods for the rapid discovery of unknown sequence differences between closely related genomes. Also, any differences found in one natural isolate could be checked in the others, providing another potentially useful shortcut toward determining whether any of the natural isolates are truly exact clones of the Ames strain and the forensic isolate. None of the several natural isolates mentioned by Read, et al.[8] proved to be exact clones of the Ames strain, although distinguishing mutations were found in the virulence plasmids (but not on the chromosome).

SOME ISSUES NEEDING FURTHER ATTENTION

In this section, we briefly raise some questions that would benefit from further attention, either in general or in specific cases that might arise.

How similar are mutation rates across species? Among strains within a species? At the level of species, published data seem to indicate that point mutation rates are inversely correlated with genome size across microbes with DNA genomes ranging all the way from viruses through bacteria to yeast.[6,15] The product of these quantities, which equals the genomic mutation rate, is surprisingly constant by this analysis. It would be useful to have more data bearing on this pattern, especially as the explanation for the genome-level constancy remains unclear. In any case, these data support the general expectation that genome-wide rates of point mutation are often well below one, which allows the existence of exact clones, especially in laboratory settings where strains may be deliberately stored and then sent elsewhere.

Despite this apparent constancy in rates across such diverse taxa, there can be substantial variation in mutation rates among different strains of the same species. For example, there are strains of *E. coli* and many other bacterial and fungal species that have mutation rates elevated by 10, 100, or even 1000-fold owing to defects in mismatch repair and other pathways.[16] Such "mutator" strains can be fairly common, both in nature[17] and in the laboratory.[12] In these mutator strains, the potential for exact clones would be greatly reduced.

The apparent constancy of genomic point-mutation rates does not extend to viruses with RNA genomes, such as the virus that causes flu in humans. RNA-based viruses, including retroviruses (such as HIV, which causes AIDS), generally have much higher genomic mutation rates despite their small genome size.[18] Again, the potential for exact clones of these viruses would be very much reduced relative to the potential in bacteria with normal DNA repair functions.

How variable are mutation rates across sites within a given genome? On the one hand, variation in mutation rates will complicate estimation of the divergence times between isolates, as well as calculation of the probability of a spurious match between isolates that converged on the same mutations. On the other hand, hypermutable sites may provide opportunities to detect sequence differences that arise quickly (between closely related isolates), and these sites can be screened for a large set of isolates at much less cost than sequencing their entire genomes.

In the case of typical point mutations, the differences between sites are probably fairly minor in the greater scheme of things. However, there are certain sequences, including homopolymeric tracts (e.g., AAAAA . . .) and VNTRs (e.g., ATTATTATT . . .) that are much more mutable, with the repeated elements being added or lost to produce insertions and deletions.[19] Examination of multiple hypermutable loci increases the chance of detecting one or more informative mutational differences even over short time scales. The *B. anthracis* genome has several such regions[20] which are recognizable and can be avoided or subjected to particular scrutiny, depending on the appropriate context.

How are strains propagated? The expected number of substitutions that distinguish two isolates will depend on the number of generations between them as well as the relevant mutation rates. The realized number of substitutions will also depend on the fate of particular mutations, and the likelihood that certain kinds of mutation are substituted will in turn depend on how a population is grown.[14] If a population experiences repeated bottlenecks, such as when cells are plated as individual colonies, then deleterious mutations can accumulate. In large populations that do not experience any severe bottlenecks, natural selection will tend to eliminate deleterious mutations and substitute beneficial ones. Neutral mutations, such as those synonymous changes that do not alter a protein's structure, will not be affected in either case. In general, because more mutations are deleterious than beneficial, the genomic substitution rate will be somewhat lower than the genomic mutation rate. However, the discrepancy should not be more than a few-fold, because onequarter or so of all possible mutations in protein-encoding genes are synonymous and therefore should be neutral or nearly so.

A related issue is the history of mutagenesis, if any, during the derivation of strains and subcultures. We have already noted that the high-temperature and antibiotic treatments used to eliminate the virulence plasmids from the Porton Down isolate of *B. anthracis* are mutagenic, and may well account for the genetic differences between that isolate and the other Ames isolates. It is generally thought that mutations do not accumulate during the long-term storage of strains in freezers or as spores, because cells are not metabolically active. However, when vegetative cells are kept at higher temperatures, they continue to metabolize and may accumulate mutations, with the extent of this starvation-induced mutability being rather variable, at least among strains of *E. coli*.[21]

CONCLUSIONS

- In contrast to human DNA-based forensics, a perfect match between a bacterial forensic sample and some potential source is less definitive owing to exact clones that may exist. Such exact clones reflect the asexual mode of bacterial reproduction.
- Microbiology laboratories often store clones and exchange them with other laboratories. The opportunity for exact clones to confound forensic source-tracking is thus especially relevant in the context of forensic samples that may have come from a laboratory.
- If a forensic sample is genetically identical to samples from several labs, then the genome sequences lack the power to discriminate between them as potential sources. However, other types of evidence (e.g., trace elements) may exist that can allow discrimination and should be sought in any case.
- Even one or a few genetic differences between a forensic sample and some potential sources can exclude certain scenarios while focusing attention on others. The pattern of differences among potential sources might, for example, implicate one of them as an intermediate or final stage in the derivation of a forensic sample.
- The issues emphasized in this chapter are more relevant for tracking a deliberate attack than a natural epidemic. In the latter case, the number of generations and accumulated genetic differences are generally much greater, the potential sources are not deliberately stored as clones, and perfect attribution is often less critical.

RECOMMENDATIONS

In closing, we offer several recommendations relevant to the issues addressed in this chapter. Some of our recommendations are similar to ones put forth in a recent report of the American Academy of Microbiology,[22] which also included others beyond the scope of our chapter. Although our recommendations are cast with reference to improving the science of microbial forensics, we emphasize that fulfilling most of them would simultaneously enhance the investigation of natural epidemics.

- The scientific community should make plans to obtain whole-genome sequences for all significant pathogenic threats to humans, ideally in advance of any forensic investigation that might become necessary. Genome sequences should also be obtained for important pathogenic threats to those plant and animal species on which we depend for food and economic vitality.
- It would be most useful to have multiple whole-genome sequences for every pathogenic species that poses a significant threat. While a single sequence would allow the extent of differences with a subsequently sequenced forensic sample to be gauged, having several diverse sequences would enable a more powerful triangulation to indicate the pattern of relatedness among strains and the order in which genetic differences accumulated.
- Basic population-genetic research in support of microbial forensics is needed to estimate rates of mutation and genomic evolution, including differences between sites in the same genome, variation among diverse species, and under alternative culture conditions. Such research should include studies in experimental settings as well as comparative studies of natural populations.
- Basic molecular biological research is needed to develop methods for the rapid discovery of unknown sequence differences between closely related genomes. Various techniques for genotyping based on known differences are well established, but methods to find, for example, a single point-mutation difference between two genomes need more research.
- Attention must be paid to statistical issues when interpreting microbial forensic evidence, as has been the case with human DNA forensics. However, the precise nature and form of the calculations will be somewhat different from those used in human forensics, owing to differences in the underlying genetic processes and population structures.

ACKNOWLEDGMENTS

We thank Chris Borland, Bruce Budowle, Charles Cooke, David Rasko, Tim Read, and Steven Salzberg for helpful discussion and correspondence. R.E.L.'s research is supported primarily by the National Science Foundation (NSF), and P.K.'s research is supported by the NSF, National Institutes of Health, Department of Energy, and the Department of Justice.

REFERENCES

1. National Research Council. *The evaluation of forensic DNA evidence*. Washington, DC: National Academy Press; 1996.
2. Li WH, Sadler LA. Low nucleotide diversity in man. *Genetics* 1991;129:513–523.
3. Cavalli-Sforza LL, Menozzi P, Piazza A. *The history and geography of human genes*. Princeton, NJ: Princeton University Press; 1994.
4. Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, Richter DJ, Lander ES, Altshuler D. Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Gen* 2002;32:135–142.
5. Chakraborty R, Kidd KK. The utility of DNA typing in forensic work. *Science* 1991;254: 1735–1739.
6. Drake JW. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Nat Acad Sci USA* 1991;88:7160–7164.
7. Lenski RE, Winkworth CL, Riley MA. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *J Mol Evol* 2003;56:498–508.
8. Read TD, Salzberg SL, Pop M, Shumway M, Umayam L, Jiang L, Holtzapple E, Busch JD, Smith KL, Schupp JM, Solomon D, Keim P, Fraser CM. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* 2002;296:2028–2033.
9. Read TD, Peterson SN, Tourasse N, Baillie LW, Paulsen IT, Nelson KE, Tettelin H, Fouts DE, Eisen JA, Gill SR, Holtzapple EK, Okstad OA, Helgason E, Rilstone J, Wu M, Kolonay JF, Beanan MJ, Dodson RJ, Brinkac LM, Gwinn M, DeBoy RT, Madpu R, Daugherty SC, Durkin AS, Haft DH, Nelson WC, Peterson JD, Pop M, Khouri HM, Radune D, Benton JL, Mahamoud Y, Jiang LX, Hance IR, Weidman JF, Berry KJ, Plaut RD, Wolf AM, Watkins KL, Nierman WC, Hazen A, Cline R, Redmond C, Thwaite JE, White O, Salzberg SL, Thomason B, Friedlander AM, Koehler TM, Hanna PC, Kolsto AB, Fraser CM. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* 2003;423:81–86.
10. Vogler AJ, Busch JD, Percy-Fine S, Tipton-Hunton C, Smith KL, Keim P. Molecular analysis of rifampin resistance in *Bacillus anthracis* and *Bacillus cereus*. *Antimicrob Agents Chemother* 2002;46:511–513.
11. Kimura M. *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press; 1983.
12. Sniegowski PD, Gerrish PJ, Lenski RE. Evolution of high mutation rates in experimental populations of *Escherichia coli*. *Nature* 1997;387:703–705.
13. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W. Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* 1978;274:775–780.
14. Elena SF, Lenski RE. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Gen* 2003;4:457–469.
15. Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics* 1998;148:1667–1686.

16. Friedberg EC, Walker GC, Siede W. *DNA repair and mutagenesis*. Washington, DC: Am Soc Microbiol Press; 1995.
17. LeClerc JE, Li BG, Payne WL, Cebula TA. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* 1996;274:1208–1211.
18. Drake JW, Holland JJ. Mutation rates among RNA viruses. *Proc Nat Acad Sci USA* 1999;96: 13910–13913.
19. Moxon ER, Rainey PB, Nowak MA, Lenski RE. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Bio* 1994;4:24–33.
20. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ, Hugh-Jones ME. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol* 2000;182:2928–2936.
21. Bjedov I, Tenaillon O, Gérard B, Souza V, Denamur E, Radman M, Taddei F, Matic I. Stress-induced mutagenesis in bacteria. *Science* 2003;300:1404–1409.
22. Keim P. Microbial forensics: A scientific assessment. Washington, DC: *Am Acad Microbiol*; 2003.

CHAPTER 17

Quality Management in Microbial Forensics Laboratories

JOSEPH M. CAMPOS

Department of Laboratory Medicine, Children's National Medical Center, Professor of Pediatrics, Pathology, and Microbiology/Tropical Medicine, George Washington University Medical Center, Washington, DC

INTRODUCTION

Criminal investigations such as the one following the Eastern U.S. anthrax exposures of 2001 depend heavily upon expert forensic laboratory support. Until recently, little information was available regarding the quality issues that should be considered when establishing a microbial forensics laboratory specializing in biocriminal activity, or when adapting an existing laboratory to perform microbial forensics work. Defining quality practices are being addressed, and guidelines will continue to evolve to effect better-quality programs. The Scientific Working Group on Microbial Genetics and Forensics (SWGMGF) was assembled in mid-2002 by the U.S. Federal Bureau of Investigation to develop the first microbial forensics quality assurance guidelines. To formulate the guidelines, the working group was assigned the following tasks:

1. Promote the development of a national microbial forensics program that is scientifically valid and rigorous,
2. Define criteria for development and validation of microbial forensics methods that will support attribution during criminal investigations, and
3. Establish national working guidelines for laboratory quality management as applied to microbial forensics.

Group members came from a diverse assortment of scientific disciplines representing several government agencies and academia.

Microbial Forensics
Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

As a clinical microbiologist member of SWGMGF, I brought to the group the experience of directing a clinical laboratory that is compliant with the statutes that are part of the Clinical Laboratory Improvement Amendments legislation of 1988 (CLIA'88). These statutes specifically exempt laboratories that are engaged solely in forensic testing from CLIA'88 oversight. The group quickly concluded that extension of CLIA'88 to include forensic testing laboratories was unfeasible and unnecessary. However, it became apparent that some of the standards to which clinical laboratories are held by CLIA'88 are relevant to forensic laboratories. We decided that by including in the relevant guidelines information derived from the CLIA'88 standards, we could improve the reliability, reproducibility, and utility of forensic test results. The guidelines are a dynamic document, and their current content can be viewed at the following websites: http://www.fbi.gov/hq/lab/fsc/backissu/oct2003/2003_10_guide01.htm and at www.science.com.

There are general areas a laboratory should address to carry out effective quality assurance and quality control practices. The SWGMGF guidelines delineate these and include sections applicable to:

— Quality assurance monitoring and the sharing of quality assurance results with laboratory staff

— Organization and management of the microbial forensics laboratory

— Personnel standards for laboratory workers and managers, including recommendations regarding competency assessment activities and continuing education requirements

— Laboratory facilities, including recommendations regarding security

— Sample control, including recommendations concerning collection, identification, storage, and chain of custody documentation of sample handling

— Validation of test procedures and revalidation of modified test procedures

— Documentation of analytical procedures to include equipment and reagent needs, step-by-step instructions, calculation of test results, interpretation and limitations of test results, quality control requirements, and pertinent literature references

— Recommendations for equipment calibration and maintenance

— Suggestions regarding the content of laboratory reports and a description of forensic testing documents that should be retained in a readily accessible format and location

— Review of laboratory reports prior to their dissemination

— Proficiency testing to evaluate the effectiveness of laboratory policies and procedures for assuring employee competence

— Procedures for rectifying errors and discrepancies in laboratory reports

— Conduction of audits of all components of the laboratory quality management program

— Preparation of a laboratory safety manual, identification of a laboratory safety officer, and institution of an on-going safety training program

— Procedures for selecting and establishing a working relationship with subcontract laboratories for provision of analytical testing services

LABORATORY QUALITY MANAGEMENT

The purpose of developing quality management guidelines for laboratories performing microbial forensic work is to promote development of testing programs that are scientifically valid and rigorous as well as to establish minimum operational criteria for application-oriented laboratories. Criteria should be developed for validation of microbial forensics methods that can be used to determine attribution during criminal investigations. Germane to this effort is the establishment of national guidelines for quality assurance and quality control that are relevant for microbial forensics laboratories.

An early step in this process is to grasp and understand the concept of laboratory quality management. An essential component of a successful quality management program is preparation of a quality manual that describes the quality systems, policies, and practices of a laboratory.

A FEW DEFINITIONS ARE IN ORDER AT THIS POINT

Attribution is the determination of the cause or source of criminal or terrorist acts.

Biocrime or *bioterrorism* is the threat or actual use of microorganisms, toxins, pests, or prions to commit criminal or terrorist acts.

Chain of custody is the tracking and documentation of the physical control of evidence.

Microbial forensics is a scientific discipline that examines microorganisms, toxins, pests, prions, and their associated ancillary products for source attribution.

Proficiency testing involves the use of materials whose identity, type, or values have been previously characterized and are used to assess the performance of a laboratory or laboratory workers to monitor performance and identify areas of the laboratory where improvements may be needed.

Quality assurance monitoring includes activities that verify that laboratory test results and associated practices are providing useful information to the consumers of the laboratory's products.

Quality control assessments include the real-time activities that validate that test conditions are conducive to yielding accurate, reproducible, and reliable laboratory test results.

Quality systems comprise the organizational structure, resources, and processes that are necessary for implementing a quality management program.

Test validation is a process through which an analytical procedure is evaluated to determine its efficacy and reliability for correctly analyzing samples or to define the limits of the assay.

There is a great deal of overlap in the activities carried out by microbial forensics and clinical laboratories. It stands to reason then that the quality management guidelines under which clinical laboratories operate should have relevance for microbial forensics laboratories and could serve as the basis for similar guidelines tailored to the needs of microbial forensics laboratories.

Clinical laboratory professionals are highly respected for the pride they have in the quality of their work. This is for good reason and supported by an effective laboratory quality management program stressed by key provisions of CLIA'88. The components of a laboratory quality management program for clinical laboratory testing were established, ratified, and became law in September, 1992. The legislation in which the program is described is known as the CLIA'88, and it has been modified by the U.S. Congress on several occasions since its original passage. In addition to defining the parameters of a laboratory quality management program, CLIA'88 also addresses other facets of clinical laboratory testing, including laboratory accreditation, validation of test methods, proficiency testing, and laboratory personnel standards. These and other aspects of laboratory testing will be discussed further later in this chapter.

LABORATORY ACCREDITATION

Accreditation of clinical laboratories conducting moderately or highly complex testing under the purview of CLIA'88 follows an on-site survey. A surveyor representing an agency or organization deemed by the Centers for Medicare and Medicaid Services conducts a biannual inspection to assess laboratory compliance with the provisions of CLIA'88. The surveyor must be granted access to all areas of the laboratory covered by the accreditation certificate. Interviews of laboratory personnel concerning compliance with the accreditation standards are conducted. The surveyor is at liberty to observe laboratory personnel performing the pre-analytic, analytic, and post-analytic phases of testing. Any or all laboratory records and data are subject to review.

The CLIA'88 statute explicitly exempts forensic testing from the oversight that governs clinical laboratory testing. One might suggest that the jurisdiction of CLIA'88 be broadened to encompass forensic testing, but such action is probably unfeasible because of the differences between clinical and forensic laboratory testing. This nature of forensic evidence samples and unique case scenarios likely warrant a separate set of guidelines. Nevertheless, some of the standards to which clinical laboratories are held are relevant to forensic laboratories. A sensible course of action would be to include those relevant standards in the forensic laboratory guidelines under development and improve the reliability and reproducibility of forensic test results.

VALIDATION OF LABORATORY TESTS

Whether clinical or forensic samples are to be tested, test procedures should be validated before being used and the results relied upon. Validation procedures should include testing samples of known content to determine the accuracy and precision of test results, the analytical sensitivity and specificity of test methods, the range over which test results are meaningful, and the reference range (normal value) for test results so that abnormal results are immediately recognized.

When previously validated test procedures are modified, revalidation of the procedures is necessary (particularly for material modifications). Use of a tentatively validated, newly developed assay should be acceptable provided it was developed to investigate a biocrime or bioterrorism event which may be necessary in an urgent manner or where a rapid analysis is needed. If such test results will be presented during courtroom testimony, test validation should be completed beforehand, or a panel of experts should be convened to confirm the validity of the procedure and determine the limitations of the results.

PROFICIENCY TESTING

The laboratory should have an ongoing mechanism for evaluating the effectiveness of its policies and procedures for assuring employee competence. Toward that end, individuals who are involved in the performance of laboratory tests should undergo proficiency testing on at least an annual basis. As much as possible, proficiency test samples should be tested in the same manner as actual forensic samples, and ideally as "blind" samples under conditions in which they are not recognized as proficiency test samples. Proficiency test samples may be obtained from external or internal sources and should be available for all test procedures performed by the laboratory. This blind proficiency test requirement may not be practical or possible for forensic cases. If not, then open blind proficiency tests should be employed. For some situations proficiency tests may not be available. In such circumstances, at least a qualifying test should be administered.

QUALITY CONTROL TESTING

The laboratory should follow established quality control (QC) procedures for monitoring and evaluating the analytical testing process of each test method. In most instances, individual test QC procedures should include assays of at least two levels of control materials on each day that testing is performed. Remedial or corrective action must be taken and documented when problems or errors are identified during QC activities. Records of QC testing should be maintained in an easily retrievable format for at least two years and perhaps longer when testing involves potentially sensitive or controversy-generating situations.

QUALITY ASSURANCE MONITORING

The goals of a laboratory's quality assurance program should include evaluating the effectiveness of the laboratory's policies and procedures, identifying and correcting problems identified, assuring the prompt reporting of test results, promising to document and investigate consumer complaints and problems, and assuring the adequacy and competency of the laboratory staff. Quality assurance activities should be selected with an eye toward addressing potential troublesome areas of the laboratory so that confirmed problems can be worked upon and resolved. Thus, quality is continuously improving. Performance improvement should be the ultimate goal. There is no point to allocating laboratory resources to gathering data that verify that a laboratory is doing well in an area in which they are known to be strong. If a quality assurance activity indicates that problems do not exist, the activity should be curtailed and another one begun. The results of quality assurance monitoring should be discussed by laboratory leaders with the laboratory staff, and the staff should be invited to participate in the identification of future activities.

COMPETENCY ASSESSMENT OF LABORATORY EMPLOYEES

A prerequisite that must be met in order to employ a staff of competent employees is the existence of up-to-date job descriptions that list the educational, training, and certification requirements as well as the duties and responsibilities of individuals occupying each job classification. The authority, responsibilities, and interrelationships of all personnel involved in the laboratory should be documented in the job description in an unambiguous fashion. Once the initial competencies of employees have been confirmed, they should be strongly encouraged to participate in regular continuing education activities in order to maintain and acquire additional competencies. Competency assessment sessions should be conducted on at least an annual basis with each employee.

The laboratory should be directed by an individual who is accountable for both the administrative and technical activities of the entire operation. A secondary tier of leaders should also be identified who are responsible for the content of laboratory reports and for providing courtroom testimony when necessary. The laboratory managerial staff should be answerable to implementing the laboratory's quality management program and for striving to achieve on-going improvement of laboratory performance.

PROCEDURE MANUALS

All laboratories should maintain a readily accessible manual containing the standard operating procedure (SOP) for each test performed. A clinical SOP should contain the necessary information regarding sample collection, preservation, transportation, and processing. Forensic SOPs may separate these into crime scene and analytical procedures. Because of the varied nature of crime scenes, evidence collection protocols should be more flexible. The SOP should list the equipment and reagents needed for testing and provide step-by-step instructions for performing the assay and calculating the test results. Also essential are descriptions of the quality control tests that should be performed, the limitations of the test results, and pertinent literature references that supply data as to the sensitivity, specificity, and predictive values of test results.

LABORATORY REPORTS

Properly configured laboratory reports include all of the information necessary for the correct interpretation of the significance of assay findings. Elements of the laboratory report that should be present include the source and a description of the sample, the date the sample was received by the laboratory, the date the testing was completed by the laboratory, the name and location of the testing laboratory, the interpretation of the test results (where interpretation is helpful), recommendations for additional testing (if appropriate), description of the analytical method(s) used, identification of components of the analysis that were performed by subcontracted laboratories, signature of the individual responsible for the content of the report, and documentation of review of the report by a competent authority.

LABORATORY RECORDS

Written records should be maintained that document all of the observations and findings made during testing. Such records may become invaluable during courtroom testimony in which allegations of poorly performed testing are being made. Laboratory records should include signed/dated procedural notes generated by the individuals performing the testing; a list of the specific equipment used during the testing; photographs, charts, and graphs related to the testing; chain of custody forms documenting the handling of the evidence by the laboratory; and information regarding the complete consumption or the disposition of remaining evidence.

LABORATORY SECURITY

The microbial forensics laboratory should have enforced policies in place regarding the physical security of the laboratory premises, protection of laboratory data and information technology, and criminal background checks on all laboratory personnel. As a general precaution, access to microbial forensic testing facilities should be controlled and limited only to individuals with a need to be admitted. Investigatory evidence and all “select agents” should be stored safely and securely when they are not actively under examination.

There should be a readily accessible safety manual available to laboratory employees that is supplemented with an annual, mandatory laboratory safety training program. Procedures for cleaning, disinfecting, and monitoring the workplace environment should be in place. One member of the laboratory should be appointed as the laboratory safety officer with the responsibility of managing the safety program.

Additional considerations for a forensic laboratory include heightened security for maintaining evidence, as well as security of data. Tampering of evidence or loss of data should not be tolerated and should be avoided to the best possible abilities of the laboratory.

CONCLUSIONS

Delivery of accurate, reproducible, and useful data should be the objective of every microbial forensics laboratory. The ability of a laboratory to achieve this objective is dependent upon many factors, not the least of which is meeting a requirement for a well-functioning quality management program. Quality is the consequence of granting unwavering attention to each of the components of the quality management program, and as described in the preceding sections of this chapter, there are a multitude of components to monitor.

Clinical laboratories in the U.S. have done a remarkable job of improving quality since the initiation of CLIA'88. Certainly the provisions in the legislation that address mandatory quality management activities are largely responsible. It is the hope of the SWGMGF that the working guidelines developed in 2003 for quality management will yield a similar outcome for the increasingly critical work being performed by microbial forensics laboratories.

BIBLIOGRAPHY

American Society of Crime Laboratory Directors/Laboratory Accreditation Board. *Accreditation Manual*, ASCLD/LAB, Garner, NC, 2003.

Code of Federal Regulations. 42CFR73. *Possession, Use, and Transfer of Select Agents and Toxins; Interim Final Rule*, Washington, DC, U.S. Government Printing Office, December 13, 2002, pp 76886–76905.

Code of Federal Regulations. 42CFR493. *Laboratory Requirements*, Washington, DC, U.S. Government Printing Office, January 1, 2003, pp 967–1087.

International Organization for Standardization. *General Requirements for the Competence of Testing and Calibration Laboratories*. ISO/IEC 17025, Geneva, Switzerland, 1999.

Murray PR (editor-in-chief). *Manual of Clinical Microbiology*, Eighth Edition, American Society for Microbiology Press, Washington, DC, 2003.

NCCLS. *Application of a Quality System Model for Laboratory Services; Approved Guideline—Second Edition*. NCCLS document GP26-A2, Wayne, PA, 2003.

NCCLS. *Assessment of Laboratory Tests When Proficiency Testing Is Not Available; Approved Guideline*. NCCLS document GP29-A, Wayne, PA, 2002.

NCCLS. *Clinical Laboratory Safety; Approved Guideline—Second Edition*. NCCLS document GP17-A2, Wayne, PA, 2004.

NCCLS. *Clinical Laboratory Technical Procedure Manuals; Approved Guideline—Fourth Edition*. NCCLS document GP02-A4, Wayne, PA, 2002.

NCCLS. *Training and Competence Assessment; Approved Guideline—Second Edition*. NCCLS document GP21-A2, Wayne, PA, 2004.

NCCLS. *Using Proficiency Testing to Improve the Clinical Laboratory; Approved Guideline*. NCCLS document GP27-A, Wayne, PA, 1999.

NIST. *Quality Assurance Standards for Forensic DNA Testing Laboratories*, Gaithersburg, MD, 1998.

U.S. Department of Health and Human Services. *Biosafety in Microbiological and Biomedical Laboratories*, Fourth Edition, U.S. Government Printing Office, Washington, DC, 1999.

CHAPTER 18

Admissibility Standards for Scientific Evidence

ROCKNE HARMON
Alameda County District Attorney's Office, Oakland, California

In the United States and abroad, much attention has been focused on scientific solutions to the new forensic problems posed by the prospect of biological weapons attacks. The threat of war and of terrorist activity at home and abroad has created a new consciousness of our vulnerability and sparked discussion of how best to respond. Deliberate release of a biological agent directed against the U.S. military or civilian population or such critical infrastructures as animal and plant agriculture and the food supply system would be either an act of war, if launched by a nation state, or a criminal act, if perpetrated by an individual or group of terrorists. The U.S. would respond by political, diplomatic, military, and legal means. For these to be most effective, it is essential to be able to investigate the event, correctly identify those responsible, and prove this in a court of law. Readers will be familiar with the concept of forensic examination of firearms and of bullets recovered from the crime scene so that the results can be presented as evidence in legal proceedings. The field of microbial forensics is developing in response to the need to examine biological agents used for criminal purposes so that the results can also be presented as evidence upon which legal, political, or military decisions may be based.

The U.S. legal system is a system of legal precedent. In such a system, to deal consistently with issues that are dealt with repeatedly, a legal precedent has been described as ". . . a legal decision that may serve as a justification for a later one."[1] For consistency, legal admissibility standards have been developed for all forms of evidence. More specific rules have been enacted for expert testimony, opinion testimony, and scientific evidence. While these rules may vary between and among jurisdictions, they all share a common goal—to ensure that whatever testimony is presented in these specialized areas is deserving of the jury's consideration. That is not to say that our legal precedent system has always dealt with the same subject consistently. Each case is

Microbial Forensics
Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

somewhat dependent on effective advocacy, legal intellect, and an adequate representation of the scientific reality.

For the new field of microbial forensics, there is much to learn from legal experiences with other forms of scientific evidence. I will begin by describing the legal admissibility standard for scientific evidence in the U.S. Federal Courts. This standard has been chosen because, in all likelihood, a bioterrorist event is likely to be prosecuted in Federal Court. This does not mean that there will not be biocrimes which are prosecuted in state courts—in fact, the human immunodeficiency virus (HIV) infection criminal case discussed below was prosecuted in the Louisiana State Court. Whether federal or state, all admissibility standards have similar goals and factors to consider. After discussing the federal admissibility standard, I will then outline legal admissibility experiences with human nuclear DNA, plant DNA, phylogenetic testing (for example, HIV), human mitochondrial DNA, and latent fingerprint evidence. Many of the examples cited occurred in state courts where other admissibility standards were utilized. Finally I will draw on prior sections to suggest how microbial forensics might smoothly transition into the legal system.

LEGAL ADMISSIBILITY

Scientific evidence has long been governed by legal admissibility standards designed to ensure that there is some level of confidence in the evidence before it may be presented in a court proceeding. While there are several different approaches to determining admissibility of scientific evidence among different U.S. jurisdictions, they all share the same philosophy, which is focused on allowing reliable evidence and excluding unreliable evidence. Because any incident requiring microbial forensics is likely to have national security implications, this chapter will primarily discuss Federal legal admissibility standards.

Until 1993, the Federal standard for admissibility was governed by a 1923 decision, *Frye v. U.S.*[2] The essence of that standard was that ". . . expert opinion based on a scientific technique is inadmissible unless the technique is *generally accepted* as reliable in the relevant scientific community." For 70 years this "general acceptance" standard was used to determine scientific evidence admissibility in Federal and many state courts. In 1975 the Federal Rules of Evidence[3] were enacted. Some of these rules specifically address foundational requirements for the admissibility of scientific evidence. From the date of enactment of the Federal rules, there was significant debate and divergence of opinion among legal scholars about whether or not the Rules superseded the general acceptance standard articulated in *Frye*. Any question about this conflict was decisively resolved in 1993 by the decision of the U.S. Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals*.[4] This opinion determined that *Frye* no longer governed the admissibility of scientific evidence, and that the Federal rules prescribed how such evidence should be scrutinized prior to admitting it in a Federal Court proceeding.

BACKGROUND TO THE *DAUBERT* DECISION

The litigation in *Daubert* arose out of allegations of birth defects allegedly sustained as a result of the mothers' ingestion of the anti-nausea prescription drug Bendectin. The challenged expert testimony involved animal studies, chemical structure analyses, and re-analysis of previously published epidemiological (human statistical) studies that supported the plaintiffs' contention that Bendectin was the cause of their children's birth defects. The District (trial) Court ruled that the proffered testimony was inadmissible under *Frye* because the methods used were not generally accepted. Most emphatically, the Court rejected the epidemiological re-analyses because they had not been subjected to peer review publication. Because this left the plaintiff without proof of the cause of the birth defects, summary judgment was granted, thereby ending the lawsuit. This decision was appealed and ultimately decided by the U.S. Supreme Court. The case was remanded back to the trial court to evaluate the proffered testimony under the Federal Rules of Evidence and utilizing the factors articulated in the opinion.

ANALYSIS OF THE *DAUBERT* DECISION

Preliminarily, *Daubert* assigns ". . . the trial judge the task of ensuring that an expert's testimony both rests on a reliable foundation and is relevant to the task at hand" (ref. 4, p. 581). The reliability and relevance criteria are explicit in Federal Rules of Evidence 702.[3] The Rule's reliability criterion is established if the expert's testimony pertains to scientific knowledge. The Rule's relevance criterion is established if the testimony will assist the trier of fact to understand the evidence or to determine a fact in issue.

In addition, the opinion identified several other factors to consider in evaluating future proffers of expert testimony. Those enumerated factors for consideration are: the testability of the science; whether or not it has been subjected to peer review; the error rate of the method; existence of standards; and whether or not the science has been generally accepted by those who are familiar with it. A brief discussion of these factors is in order.

TESTABILITY

In this context, the *Daubert* opinion discusses ". . . whether or not it can be (or has been) tested (p. 592)," that it ". . . must be capable of empirical test

(p. 592)," and ". . . its falsifiablility, or refutability, or testability p. 592." The opinion also suggested that the science ". . . must be supported by appropriate validation. . . ." (ref. 4, p. 591). The reasons for these requirements are because ". . . evidentiary reliability will be based on scientific validity" (ref. 4, p. 591). It is clear from the opinion that this factor describes the scientific method—only allowing hypotheses which have been tested and proven.

PEER REVIEW

The *Daubert* opinion recognized that publication is but one element of peer review, but that ". . . submission to the scrutiny of the scientific community is a component of good science, in part because it increases the likelihood that substantive flaws in the methodology will be detected" (ref. 4, p. 593). The fact of publication is described as relevant, though not dispositive. The factor also recognizes that some passage of time after publication is appropriate.

ERROR RATE/STANDARDS

The opinion simply states that ". . . the court ordinarily should consider the known or potential rate of error" (ref. 4, p. 594). The only authorities cited for this factor were legal opinions relating to spectrographic voice identification techniques. In both of the opinions cited, testimony was presented that demonstrated empirically determined false elimination and identification rates. At first glance it appears that the error rate in question is the error rate for the methodology. However, given the degree of subjectivity inherent in voice identification technology, it is clear that the error rate provided in those cases includes some component of operator proficiency. In any event, this factor presumably will be satisfied by a good faith effort to quantify the error rate, if possible, to allow the trier of fact to integrate that information into decision-making. The standards consideration alludes to standards created by a representative industry or community group.

This factor should not be confused with the certainty with which a conclusion is expressed or with the magnitude of the quantitative expression of the result. The *Daubert* opinion acknowledges that ". . . it would be unreasonable to conclude that the subject of scientific testimony must be "known" to a certainty. . . ." (ref. 4, p. 591).

GENERAL ACCEPTANCE

"General acceptance" is described in *Daubert* as ". . . that austere standard. . . ." (ref. 4, p. 590). Decades of legal precedent describe general acceptance as the acceptance of a technology by a majority of those who are familiar with it. Implicit is that the technology has been peer reviewed and available to the appropriate community for some time. Previously this was the exclusive factor. Now, under *Daubert*, it may not even play much of a role in the court's consideration. (The brief discussion of latent fingerprints below demonstrates this point.)

In a subsequent decision, *Kumho Tire Co., Ltd. v. Carmichael*,[5] the Supreme Court further explained ". . . that a trial court may consider one or more of the more specific factors that *Daubert* mentioned when doing so will help determine that testimony's reliability. But, as the Court stated in *Daubert*, the test of reliability is flexible, and *Daubert's* list of specific factors neither necessarily nor exclusively applies to all experts or in every case" (ref. 5, p. 563).

As a result of the U.S. Supreme Court's decision in *Daubert*, the case was remanded to the 9th Circuit Court of Appeal for reconsideration of the challenged proffered expert testimony using the factors described in *Daubert* to reevaluate the admissibility of the testimony. Using the new standard, the 9th Circuit again sustained the trial court's ruling barring the expert testimony in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[6] First the Court expressed concern that the plaintiff's experts' opinions did not reflect the consensus of opinion on the respective subjects. The Court then proposed that, "One very significant fact to be considered is whether the experts are proposing to testify about matters growing naturally and directly out of research they have conducted independent of the litigation, or whether they have developed their opinions expressly for purposes of testifying. That an expert testifies for money does not necessarily cast doubt on the reliability of his testimony, as few experts appear in court merely as an eleemosynary gesture. But in determining whether proposed expert testimony amounts to good science, we may not ignore the fact that a scientist's normal workplace is the lab or the field, not the courtroom or the lawyer's office" (ref. 6, p. 1317). In footnote 5 to this paragraph, the Court explained that, "There are, of course, exceptions. Fingerprint analysis, voice recognition, DNA fingerprinting and a variety of other scientific endeavors closely tied to law enforcement may indeed have the courtroom as a principal theatre of operations. *See, e.g., United States v. Chischilly*,[7] (admitting expert testimony concerning a DNA match as proof the defendant committed sexual abuse and murder). As to such disciplines, the fact that the expert has developed an expertise principally for purposes of litigation will obviously not be a substantial consideration" (ref. 6, p. 1317).

In the context of microbial forensics, this observation will have a profound impact in a jurisdiction using the *Daubert* admissibility standard. First, almost all research done to date was done ". . . naturally and directly . . . independent of the litigation" (ref. 6, p. 1317). Additionally, the footnote to this discussion seems to create a public sector societal benefit exception that recognizes that altruistic motives should prevail over the financial incentive that necessarily flows from civil lawsuits.

THE DNA ADMISSIBILITY LITIGATION

In the late 1980s, forensic DNA typing services slowly made their way into the criminal justice system (*Andrews v. State*).[8] At that time, the primary sources of these services were a few private labs. Initial attempts to introduce forensic DNA evidence were successful in all jurisdictions regardless of the form of legal admissibility standard. As a result of a multidisciplinary meeting at Cold Spring Harbor's Banbury Center in 1988,[9] questions were raised about certain technical and population genetic aspects of the DNA typing then available. In 1992, a report of the National Research Council[10] threw a monkey wrench into the orderly introduction of forensic DNA evidence into the criminal justice system. This highly controversial report questioned whether specific issues about the population match frequency estimates had been adequately addressed. The report recommended that an artificially conservative calculation be utilized until those specific issues were addressed. The report had a short-term impact in some jurisdictions. Appellate courts occasionally ruled that DNA evidence in some cases had been improperly admitted without the conservative calculation method (see, e.g., *People v. Barney*[11]). Because of the furor spurred by its 1992 report, the National Research Council established a new committee to review these issues. Finally, in 1996, a report issued by the National Academy of Science Press[12] completely quelled the disturbance to the legal admissibility progress caused by the 1992 report. Since 1996, newer forms of human, animal, and plant DNA analyses were admitted without significant admissibility litigation, as described below.

THE PALO VERDE TREE PODS

After several years of experience with human forensic DNA typing, a criminal investigation and prosecution in Arizona developed a key piece of evidence in the form of tree seed pods found in the suspect's truck bed. These pods were to be compared with pods from a tree near the victim's body that showed signs of fresh damage. There was significant other evidence connecting the suspect to both the victim and the location where her body was found. The suspect acknowledged contact with the victim but denied having been in the area where her body was found ". . . for years. . . ." (*State v. Bogan*[13]).

The prosecution employed a local professor of plant molecular genetics who used randomly amplified plant DNA (RAPD) analyses on the evidence pods, and numerous other pods systematically collected from a "representative" sampling of trees near the victim's body and from more distant areas. Because of an existing legal opinion derived from forensic human DNA typing, the expert was not allowed to present his statistical calculation about the relative rarity of the matching pod DNA samples. He was allowed to present his opinion that he was confident that the seed pod DNA from the accused's truck matched the pod DNA from the damaged tree near the victim's body. During the admissibility hearing a defense expert challenged the match frequency, but not the fact of the match under the *Frye* admissibility standard. The fact that RAPD had never been used in any forensic context, plant or otherwise, did not weigh against the fact that RAPD had been generally accepted among those plant biologists who were familiar with it.

THE FLORIDA DENTIST CONTROVERSY

The story of Kimberly Bergalis, a young woman who became infected with HIV, can be very instructive to scientists who suddenly find themselves and their science involved in high-profile legal/scientific disputes. Ms. Bergalis had two of her wisdom teeth removed in 1987. She died in December 1991. The alleged source of her infection was her admittedly homosexual dentist, David Acer, who was infected and died in September 1990.[14] Initial speculation was that the dentist's drill handpiece was the means of transmission, because Acer himself had been treated with the drill when he was already infected. One university researcher even said that there was ". . . no doubt . . ." that the hand piece infected Bergalis and several other patients.[15] The results of epidemiological analyses commissioned by the Centers for Disease Control and Prevention (CDC) were published in 1992.[16] These results pointed to Acer as the source of the infection, based on statistical analyses comparing similarities between strains isolated from those involved. Other prominent scientists raised questions about the statistical methodologies and conclusions drawn from them.[17] The issue was revisited and the results again supported CDC's conclusion that Acer was the likely source of Bergalis' infection.[18] The incident was given the full media treatment by *60 Minutes*.[19] Much of the episode dealt with the victim's lifestyle information dug up by defense investigators that identified possible HIV risk factors having nothing to do with Acer. The Associate Producer of the segment also detailed his arguments and challenges in separate articles.[20,20a]

The Bergalis lawsuit was settled for $1 million in December 1991.[21] Other lawsuits were similarly resolved. Because the case was settled, the epidemiological and statistical evidence was never subjected to any legal admissibility challenge. (Florida's legal admissibility standard is somewhat different than that of *Daubert*, but the goal is the same—to screen out unreliable scientific evidence.) However, the case does serve as an example of how seemingly well-intentioned scientists may disagree about methodologies or conclusions once the discussion is publicized and sensationalized. It is also easy to see how conflicts like those manifest in the general literature on this subject could result in contentious admissibility challenges.

Similar phylogenetic techniques have been used as evidence in at least one criminal prosecution since the *Bergalis v. Acer* civil suit. In a 1994 criminal homicide prosecution, phylogenetic tree analysis results were ruled admissible as evidence in a Louisiana state court relying on an admissibility analysis utilizing the *Daubert* standard (*State v. Schmidt*[22]). The defendant, a physician, was charged with deliberately infecting his paramour with HIV-tainted blood. The comparison was between the HIV strain of the victim and a strain from one of the defendant's patients. Phylogenetic analysis showed that the strains were "closely related." An extensive admissibility hearing was held involving numerous highly qualified experts called by each side. The issues were reminiscent of those that polarized the debate in the *Bergalis v. Acer* saga. The judge ruled that the evidence should be admitted, and the defendant was convicted of murder based on this and other evidence.

HUMAN MITOCHONDRIAL DNA

In some situations, human mitochondrial DNA may be the only form of forensic typing possible. Mitochondrial sequencing of the control region can provide significant information which may assist in resolving issues in a criminal case, even though the frequency information provided is substantially less discriminating than with other forms of DNA typing. Other technical issues have arisen which also serve to limit the amount of discriminating information provided by this approach. For example, it has been documented that different samples from the same person may produce sequences that differ slightly.[23] To accommodate this phenomenon, interpretation guidelines have been developed to ensure that courts can have confidence in results produced by this technology.[24]

While the profile frequency is much less discriminating with mitochondrial DNA than that produced by other forms of forensic DNA typing, in the context of any investigative scenario the results may nevertheless be as persuasive to a jury as a profile frequency match between samples of one in a trillion. The following summary of a murder case illustrates this point. The four-year-old female victim was left in the care of a friend while her mother made the rounds of local bars. She was found lifeless early the next morning: the cause of death was asphyxiation. In addition, she had been raped and sodomized. No biological evidence of sexual assault was found in or on the victim's body. During the autopsy, a foreign pubic hair was found in the victim's pharynx and subjected to mitochondrial DNA typing. The hair mitochondrial DNA sequence compared favorably to defendant Ware's known sequence. That sequence had never been observed in any of the 742 samples previously typed and catalogued in the Federal Bureau of Investigation's (FBI's) database (*State v. Ware*[25]). In the context of the rest of the evidence against Ware, the mitochondrial DNA evidence helped convict him of first-degree murder.

More recently, another court reviewed the admissibility of human mitochondrial DNA; *Magaletti v. State*[26] involved a murder prosecution in Florida. Defendant's fingerprints were found on the inside of the victim's door underneath the door knob. Microscopic and mitochondrial analysis of a hair adhering to bindings on the victim associated the hair to the defendant, not the victim. In upholding the trial court's decision to admit the DNA typing results, the Court attached significance to the fact that ". . . the counting method is the *only* method of reporting used by analysts in the United States." (*Magaletti v. State*, p. 526) This statement recognizes the importance of reaching consensus for the admissibility of scientific testimony. Florida is a *Frye* jurisdiction, but this decision relied on decisions from other states, some of which utilize the *Daubert* standard.

FINGERPRINT ADMISSIBILITY LITIGATION UNDER *DAUBERT*

Although the U.S. Supreme Court explained in its shift from *Frye* to the Federal Rules of Evidence that ". . . a rigid general acceptance requirement would be at odds with the liberal thrust of the Federal Rules and their general approach of relaxing the traditional barriers to opinion testimony" (ref. 4, p. 589), the application of *Daubert* has not worked so smoothly. Fingerprint identification has been used in the legal system since the turn of the 20th century. To say that it has been generally accepted would be a gross understatement. In any jurisdiction using *Frye* as an admissibility standard, no legal challenge would be allowed because of the universal acceptance of fingerprints. With the *Daubert* decision and relegation of "general acceptance" to only one of several possible admissibility criteria, there have been numerous legal challenges to fingerprint testimony in Federal Courts.

The challengers have conceded that fingerprint examiners have generally accepted that there are procedures available to compare known and unknown fingerprints, but they took issue with the testability, peer review, and error rate/standards components of the fingerprint identification methods (*United States v. Plaza*[27]). Extensive admissibility hearings have been held in some cases. Other cases have relied on the hearing records of cases previously conducted and decided. In all cases, fingerprint identification testimony was ultimately allowed to be presented to the jury. However, because general acceptance is not a deciding factor as it was under *Frye*, there is nothing to necessarily preclude having an admissibility hearing challenging the evidence. This is a major distinction between *Frye* and *Daubert*: in a *Frye* jurisdiction, no further admissibility challenge is allowed until the consensus about general acceptance has changed. In a *Daubert* jurisdiction, the possibility of an admissibility hearing always exists, even for evidence of longstanding utility, community acceptance, and proven reliability.

OBSERVATIONS

What can be learned from these experiences? The DNA litigation was intense and expensive for a period of time. Human nuclear DNA typing was capable of producing near unique identification. The services were often provided by private labs. In the beginning, there were no organizations creating guidelines, little peer-reviewed literature, and little forensic experience. Except for a few highly publicized exceptions, forensic DNA typing cleared most admissibility standards and was used to convict innumerable guilty defendants. Ultimately, groups such as the Scientific Working Group for DNA Analysis Methods (SWGDAM) and the DNA Advisory Board (www.cstl.nist.gov/biotech/strbase/dabqas.htm) were formed to assist and oversee the many labs providing forensic DNA services. These groups have been a significant factor in ensuring that newer forms of forensic DNA will pass legal admissibility muster. The creation of the Scientific Working Group for Microbial Genetic Forensics (SWGMGF)[28] by the FBI is intended to serve similar purposes.

The Palo Verde tree pod litigation demonstrated that a legal admissibility standard can be satisfied if an existing, generally accepted method (RAPD) is used on a species to which it has never previously been applied. In addition, as long as scientific rigor and care are used to describe the significance of the match, the admissibility standard will be satisfied, even if there is a legitimate dispute about how best to express that significance. This evidence was deemed admissible without requiring that there be a Scientific Working Group or similar organization publishing standards or guidelines. The Florida Dentist litigation demonstrates very similar points.

Human mitochondrial DNA legal admissibility experience demonstrates how the combined influence of SWGDAM, extensive peer-reviewed publications, and acknowledgment of inherent anomalies in the typing system have all contributed to near-universal legal admissibility acceptance. Finally, the latent fingerprint admissibility litigation demonstrates a fundamental difference between the *Daubert* approach and *Frye*. In Federal Court there is always the opportunity to launch a legal admissibility challenge, even for evidence such as fingerprints that have been generally accepted for decades.

While factors such as peer review, general acceptance, accreditation, or the imprimatur of a Scientific Working Group are not necessarily prerequisites to a favorable legal admissibility determination, they do provide a solid basis for a court's determination that there is enough confidence in the proffered evidence to allow a jury to hear it. There may be instances where, for a variety of reasons, these indices of reliability are not present. Even then, the scientific evidence may be of great assistance to investigators in developing a case. Information in this category is sometimes described as an "investigative lead." In an ideal world, all investigative leads consist of admissible evidence and will eventually be presented as evidence in a jury trial. On the other hand there is no requirement that only admissible evidence be used for investigative purposes. One obvious example illustrates the principle. In most jurisdictions, evidence of polygraph test results is legally inadmissible by statute. However, in many investigations polygraph results are relied upon by experienced investigators in developing leads and obtaining search or arrest warrants. As long as the investigative leads produce enough evidence to sustain a conviction, there is no requirement that the basis for the leads themselves be admissible.

In summary, at this point the field of microbial forensics has made great strides in anticipation of bioterrorist or biocrime events. The establishment of SWGMGF has created the mechanism by which community standards will be established and promulgated.

REFERENCES

1. *Black's Law Dictionary*, 8th ed. Minnesota: Thomson West Publishing Company.
2. *Frye v. U.S.* 293 F. 1013 (D.C. Cir. 1923).
3. Federal Rules of Evidence. 28 U.S. Code Annotate.
4. *Daubert v. Merrell Dow Pharmaceuticals* 509 U.S. 579 (1993).
5. *Kumho Tire Co., Ltd. v. Carmichael* 526 U.S. 137 (1999).
6. *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 43 F.3d 1311 (9th Cir. 1995).
7. *United States v. Chischilly,* 30 F.3d 1144, 1153 (9th Cir. 1994).
8. *Andrews v. State* 533 So.2d 841 (Fla. App. 1988).
9. DNA Technology and Forensic Science, 32 Banbury Report, Cold Spring Harbor Laboratory Press, 1989.
10. DNA Technology in Forensic Science, National Research Council, National Academy of Science Press, 1992.
11. *People v. Barney*, 8 Cal. App. 4th 798 (1992).

12. The Evaluation of Forensic DNA Evidence, National Academy of Science Press, 1996.
13. *State v. Bogan*, 905 P.2d 515 (Ariz. App. 1995).
14. "What If the Dentist Didn't Do It?" *The New York Times*, April 16, 1994.
15. "Dentist May Have Spread AIDS With Part of Drill," *The Miami Herald*, July 13, 1991.
16. Ou CY, Ciesielski CA, Myers G, Bandea CI, Luo CC, Korber BT, et al. Molecular epidemiology of HIV transmission in a dental practice. *Science* 1992;256:1165–1171.
17. DeBry R, Abele LG, Weiss SH, Hill MD, Bouzas M, Lorenzo E, et al. Dental HIV transmission? [Letter]. *Nature* 1993;361:691.
18. Crandall K. Intraspecific phylogenetics: Support for dental transmission of human immunodeficiency virus. *J Virol* 1995;69:2351–2356.
19. "60 Minutes." Transcripts, v.XXVI, no. 38:3.
20. Barr S. 1996. The 1990 Florida dental investigation: Is the case really closed? *Ann Intern Med* 124:250–254.

20a. Barr S. In Defense of the AIDS dentist. *Lear's Magazine*, 1994; April:68–82.

21. See www.wpblawyers.com/specialties.htm.
22. State v. Schmidt 699 So.2d 448 (1997 La. App.).
23. Bendall KF, Sykes BC. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am J Hum Genet* 1995;7:248–256.
24. Budowle B, DiZinno JA, Wilson MR. Interpretation guidelines for mitochondrial DNA sequencing, 1999. In Proceedings of the Tenth International Symposium on Human Identification.
25. *State v. Ware*, 1999 WL 233592 (Tenn. Crim. App.).
26. *Magaletti v. State*, 2003 WL 1786081 (Fla. App.).
27. *United States v. Plaza*, 188 F.Supp.2d 549 (2002 E.D.Pa.).
28. Budowle B. Defining a New Forensic Discipline: Microbial Forensics. In *Profiles in DNA*, March 2003;7–10.

Quality Assurance Guidelines for Laboratories Performing Microbial Forensic Work

PRODUCED BY THE FEDERAL BUREAU OF INVESTIGATION-HOSTED MEMBERS OF THE SCIENTIFIC WORKING GROUP ON MICROBIAL GENETICS AND FORENSICS (SWGMGF)

June 20, 2003

PREFACE

The Scientific Working Group on Microbial Genetics and Forensics (SWGMGF) has developed the following quality assurance guidelines to provide laboratories engaged in microbial forensic analysis with a framework to implement a quality assurance program. This document provides guidance to laboratories that carry out microbial forensic analyses to support the judicial system. Consideration may be given to alternate methods of achieving the intent of these quality assurance practices as outlined in these guidelines. A quality program is always evolving, and likewise, this document should be considered a living document. It is the intent of SWGMGF to modify this document as quality assurance practices in microbial forensics advance.

INTRODUCTION

The *Quality Assurance Guidelines for Laboratories Performing Microbial Forensic Work* provide a baseline from which laboratories may structure their quality assurance practices. Many of the recommendations outlined in this document will be familiar to laboratories already meeting ISO 17025 and CLIA'88 requirements. It is the intent of SWGMGF to provide the community with a source document for quality practices in the microbial forensics field.

Microbial Forensics
Copyright © 2005 by Elsevier Inc. All rights of reproduction in any form reserved.

1. SCOPE

These guidelines describe quality assurance activities that a laboratory should follow to ensure the competency of the laboratory and the quality and integrity of scientific data. A laboratory, in the context of these guidelines, is defined as a facility in which microbial forensic testing is performed. These guidelines do not preclude the participation of a laboratory, by itself or in collaboration with others, from participating in research and development, or from using procedures that have not yet been validated by standard means for purposes other than forensic testing.

2. DEFINITIONS

As used in these guidelines, the following terms have the meanings specified:

Administrative review is an evaluation of examination documentation for consistency with laboratory policies and for editorial correctness.

Analytical procedure is an orderly step-by-step procedure designed to ensure operational uniformity and to minimize analytical drift.

Attribution is the information obtained regarding the identification or source of a material to the degree that it can be ascertained.

Audit is an inspection used to evaluate, confirm, or verify activities related to quality.

Biocrime or **bioterrorism** is the threat or use of microorganisms, toxins, pests, prions, or their associated ancillary products to commit acts of crime or terror.

BMBL is the Biosafety in Microbiological and Biomedical Laboratories (see Bibliography).

Calibration is a set of operations that establish, under specified conditions, the relationship between values provided by a measuring instrument, measuring system, and a known material or known values.

CDC is the Centers for Disease Control and Prevention.

Chain of custody is the tracking and documentation of physical control of evidence.

Contributing agency is identified as the organization that submitted the evidence to the laboratory for testing and which will receive the laboratory report.

Control samples or known samples are test materials whose identity, type, or values have been established (e.g., blind samples, negative and positive controls).

Critical equipment or instruments are those requiring calibration prior to their initial use and on a regular basis thereafter.

Critical reagents are determined by empirical studies or routine practice to require testing on known samples prior to use with evidentiary materials in order to prevent unnecessary consumption of forensic samples.

Derivative evidence is material having originated from the original evidence (e.g., grown cultures, amplified DNA).

Examination documentation encompasses any documentation generated as a result of the analysis of submitted evidence. This may include technical notes, worksheets, charts, graphs, printouts, spectra, photographs, and other data or records used by examiners/analysts to support their reported conclusions.

Examiner/analyst (or equivalent role, position, or title as designated by the laboratory director) is an individual who conducts and/or directs the analysis of samples, interprets data, and reaches conclusions, and may eventually testify to those findings or conclusions.

Expert testimony is an opinion rendered by a qualified individual based on technical expertise or training.

Investigative support is data that provide the contributing agency with basic information to assist an ongoing investigation.

Laboratory is a facility in which microbial forensic testing is performed, a facility that contracts with another entity for such testing, or an entity contracted to perform analyses.

Laboratory support personnel (or equivalent role, position, or title as designated by the laboratory director) is an individual(s) performing supportive laboratory duties, but is not involved in the analysis or interpretation of evidence.

Microbial forensics is a scientific discipline that examines microorganisms, toxins, pests, prions, or their associated ancillary products for source attribution.

Proficiency test samples are materials whose identity, type, or values have been previously characterized and are used to assess the performance of a laboratory or an individual(s).

Proficiency testing is a quality assurance measure used to monitor performance and identify areas where improvements may be needed. Proficiency tests may be classified as:

- Internal: A proficiency test that is prepared and administered within the laboratory being tested.
- External: A proficiency test that is prepared by an external agency.

Qualifying test measures an individual's proficiency in both technical skills and knowledge and is administered prior to an individual assuming independent work responsibility.

Quality assurance is the system of management activities designed to ensure that a process, item, or service is of the type and quality needed. This includes monitoring activities that are intended to verify whether practices and test results are providing reliable and relevant information.

Quality control is a mechanism or laboratory activity intended to verify whether test conditions are functioning appropriately to yield reproducible results.

Quality manual is a document stating the quality policy, quality system, and quality practices of an organization.

Quality system is the organizational structure, responsibilities, procedures, processes, and resources for implementing quality management.

Reference material (certified or standard) is a material for which identities, types, or values are certified by technically valid procedures and is accompanied by, or traceable to, a certificate or other documentation.

Secure area is a locked space (e.g., cabinet, vault, room) with access restricted to authorized personnel.

Subcontractor is an individual or entity having a transactional relationship with a laboratory.

Technical manager (or equivalent position or title as designated by the laboratory director) is the individual who is accountable for the technical operations of the laboratory.

Technical review is an evaluation of reports, notes, data, and other documents to ensure that an appropriate and sufficient basis exists for the reported scientific conclusions.

Technician (or equivalent role, position, or title as designated by the laboratory director) is an individual who performs analytical techniques on samples under the supervision of an examiner/analyst.

Traceability is the property of a result of a measurement whereby it can be related to appropriate standards, generally international or national standards, through an unbroken chain of comparisons.

USDA is the United States Department of Agriculture.

Validation is a process by which a procedure is evaluated to determine its efficacy and reliability for analysis and includes the following:

- Developmental validation is the acquisition of test data and the determination of conditions and limitations of a newly developed methodology for use on samples.
- Preliminary validation is the acquisition of limited test data to enable an evaluation of a method used to assess materials derived from a biocrime or bioterrorism event.
- Internal validation is an accumulation of test data within the laboratory to demonstrate that established methods and procedures perform within determined limits in the laboratory.

3. QUALITY ASSURANCE PROGRAM

3.1. The Laboratory Should Establish and Maintain a Documented Quality System That Is Appropriate to the Testing Activities

3.1.1. The Quality Manual Should Address, at a Minimum, the Following

- Goals and objectives
- Organization and management
- Personnel qualifications and training
- Facilities and security
- Sample control
- Validation
- Analytical procedures
- Calibration and maintenance
- Proficiency testing
- Corrective action
- Documentation and report writing
- Review of reports
- Safety
- Audits
- Monitoring subcontractor work, if a subcontractor is used

3.2. The Laboratory Should Identify An Individual As the Quality Manager

4. ORGANIZATION AND MANAGEMENT

4.1. A Laboratory Should

- Have a managerial staff with the authority and resources needed to discharge their duties and meet the criteria of the guidelines in this document.
- Have a technical manager who is accountable for the technical operations.
- Specify and document the responsibility, authority, and interrelation of all personnel who manage, perform, or verify work affecting the validity of microbial forensic analysis.

4.2. A Laboratory Should Have a Procedure for Document Control

5. PERSONNEL

5.1. Laboratory Personnel Should Have the Education, Training, and Experience Necessary to Perform Examinations and Provide Testimony. The Laboratory Should

5.1.1. Have Written Job Description(s) for Personnel that Include Responsibilities, Duties, and Required Skills

5.1.2. Have a Documented Training Program for Assuring the Competence of All Technical Laboratory Personnel

5.1.3. Have a Documented Program to Ensure Technical Qualifications Are Maintained Through Continuing Education

5.1.3.1. Continuing Education—The Technical Manager and Examiner/Analyst(s) Should Stay Abreast of Developments within the Field of Microbial Forensics by Reading Current Scientific Literature and by Attending Seminars, Courses, Professional Meetings, or Documented Training Sessions/Classes in Relevant Subject Areas at Least Once a Year

5.1.4. Maintain Records on the Relevant Qualifications (Including College Transcripts), Training, Skills, and Experience of All Technical Personnel

5.1.5. Maintain Records of Security Clearances, When Appropriate, for Laboratory Personnel, When Evidence/Information Must be Secure and Protected

5.1.6. Obtain a Nondisclosure Agreement from Employee, If Required

5.2. The Technical Manager Should Have the Following

5.2.1. Degree/Experience Requirements: The Technical Manager of a Laboratory Should Have, at a Minimum, One of the Following Combinations of Education/Experience

- Master's degree in biology, microbiology, clinical laboratory science, or other relevant science. Within the last eight years: a minimum of four years of work experience in a laboratory with at least one year in the relevant field and one year of supervisory experience.
- Doctorate degree in medicine, biology, microbiology, clinical laboratory science, or other relevant science. Within the last four years: a minimum of two years' work experience in a laboratory with at least one year in the relevant field and one year of supervisory experience.

5.2.2. Duty Requirements

5.2.2.1. General: Manages the Technical Operations of the Laboratory

5.2.2.2. Specific Duties

- Is responsible for evaluating and approving all methods used by the laboratory and for proposing new or modified analytical procedures to be used by examiners.
- Is responsible for technical problem solving of analytical methods and for the oversight of training, quality assurance, safety, and proficiency testing in the laboratory.
- Has the authority to suspend laboratory operations.

5.2.2.3. The Technical Manager Should be Accessible to the Laboratory to Provide On-Site, Telephonic, or Electronic Consultation, As Needed. The Technical Manager Should Spend at Least 50% of the Work Year at the Laboratory Facility

5.3. Examiner/Analyst Should Have the Following

5.3.1. Degree Requirements: An Examiner/Analyst Should Have, at a Minimum, a Bachelor's Degree or Its Equivalent in Biology, Microbiology, Clinical Laboratory Science, or Other Relevant Science

5.3.2. Experience Requirements: An Examiner/Analyst Should Have a Minimum of 12 Months of Laboratory Experience

5.3.3. Duty Requirements: An Examiner/Analyst Is Responsible for the Content of a Laboratory Report Released Under His/Her Name and Provide Expert Courtroom Testimony, If Required

5.3.4. Training Requirements: An Examiner/Analyst Should Have Successfully Completed a Documented Training Program and Qualifying Test Before Assuming Independent Work Responsibilities. The Training Program Should Require a Minimum of Four Months and Include the Successful Analysis of a Range of Samples Typically Encountered in His/Her Area of Microbial Forensics and Provide an Understanding of Forensic Operations

5.4. Technician Should Have

5.4.1. On-the-Job Training Specific to the Job Function(s) Performed

5.4.2. Successfully Completed a Qualifying Test Before Participating in Forensic Testing Responsibilities. Documentation of Testing Should be Retained

5.5. Laboratory Support Personnel Should Have Training, Education, and Experience Commensurate with Their Responsibilities as Outlined in Their Job Description

6. FACILITIES

6.1. The Laboratory Should Have a Facility That Is Designed to Provide Appropriate Levels of Security, Safety, and Minimize Contamination. Laboratory Management Should Ensure That

6.1.1. Access to the Laboratory is Controlled and Limited

6.1.2. Evidence is Appropriately Secured and Stored When Not Under Examination

6.1.3. Preamplification Materials and Activities are Separated by Time or Space from Postamplification Products and Activities

6.1.4. The Laboratory Follows Written Procedures for Monitoring, Cleaning, and Decontaminating Facilities and Equipment

6.1.5. When CDC (BMBL) and/or USDA Guidelines for Biolevel Containment are Employed, Appropriate Protocols and Records Should be Maintained

6.2. A Laboratory Should Document Waste Management Requirements to Ensure That Appropriate Decontamination and Disposal Measures Are in Accordance with Federal and/or State Law

6.3. A Laboratory Should Consider Measures to Address the Following

- Backup power sources for power outages to protect evidence from temperature fluctuations, when appropriate. Twenty-four-hour emergency notification procedures are advisable.
- Computer security and backups for data and case documentation records.

7. SAMPLE CONTROL

7.1. The Laboratory Should Have and Follow a Documented Sample Inventory Control System. This System Should Ensure That

7.1.1. Samples are Marked with Unique Identifiers

7.1.2. Documentation of Sample Identity, Receipt, Storage, and Disposition is Maintained

7.1.3. The Laboratory Follows Documented Procedures That Minimize Sample Loss, Contamination, and/or Deleterious Change

7.1.4. The Laboratory Has Secure Areas for Sample Storage Including Environmental Control Consistent with the Form or Nature of the Sample

7.2. The Laboratory Should Have Available Guidelines for Sample Collection and Should Request the Following Information from the Submitter of Samples

7.2.1. Sample Collected from an Individual: If Available, Type of Sample (Method of Collection and Anatomical Site from Which Collected), Age, Gender, Population Affinity, Weight, Body Temperature, Current Medications, Who Collected Sample, Time and Date of Collection, and Post-Collection and Transportation Conditions. When Multiple Samples are Collected from the Same Individual, Appropriate Delineation Should Be Provided to Distinguish Samples. Additionally, for Nonhuman Samples, Species, Animal, or Plant Identification Number Should Be Recorded

7.2.2. Sample Collected from a Location: If Available, Type of Sample, Sampling Tool, Date and Time Collected, and Who Collected the Sample. When Multiple Samples are Collected from the Same Site, Appropriate Delineation Should Be Provided to Distinguish Samples. Sample Location and Condition of Location at Time of Sample Collection Should Be Recorded. Conditions of Storage and Transport Should Be Documented

7.3. The Laboratory Should Maintain a Chain of Custody for Forensic Samples From the Time of Receipt in the Laboratory. Individual Items Should Be Tracked. Time and Date of Sample Transfers Should Be Documented. Derivative Evidence Should Be Tracked

7.4. When the Laboratory Consumes Tested Samples, That Fact Should Be Documented in the Case Notes

7.5. If Tested Samples, or Derivative Evidence Thereof, Are Released to a Database, This Should Be Recorded in the Case Notes and the Chain-of-Custody

7.6. A Laboratory Should Document and Provide Appropriate Guidelines for Sample Submission, Packaging, and Return

7.7. A Laboratory Should Document a Policy of Long-Term Sample Storage, Retention, Disposal, and/or Return

8. VALIDATION

8.1. The Laboratory Should Use Validated Methods and Procedures for Analyses

8.1.1. Developmental Validation Should Be Appropriately Documented and Should Address Specificity, Sensitivity, Reproducibility, Bias, Precision, False-Positives, False-Negatives, and Determine Appropriate Controls. Any Reference Database Used Should Be Documented

8.1.2. Preliminary Validation Is the Acquisition of Limited Test Data to Enable an Evaluation of a Method Used to Provide Investigative Support to Investigate a Biocrime or Bioterrorism Event. If the Results Are to Be Used for Other Than Investigative Support, Then a Panel of Peer Experts, External to the Laboratory, Should Be Convened to Assess the Utility of the Method and to Define the Limits of Interpretation and Conclusions Drawn

8.1.3. Internal Validation Should Be Performed and Documented by the Laboratory

8.1.3.1. The Procedure Should Be Tested Using Known Samples. The Laboratory Should Monitor and Document the Reproducibility and Precision and Define Reportable Ranges of the Procedure Using Control(s)

8.1.3.2. Before the Introduction of a New Procedure into Sample Analysis, the Analyst or Examination Team Should Successfully Complete a Qualifying Test for That Procedure

8.1.3.3. Material Modifications Made to Analytical Procedures Should Be Documented and Subjected to Validation Testing Commensurate with the Modification and Have Documented Approval

9. ANALYTICAL PROCEDURES

9.1. The Laboratory Should Have and Follow Written Analytical Procedures Reviewed and Approved by the Laboratory Management/Technical Manager

9.1.1. The Laboratory Should Have a Documented Procedure for Each Analytical Technique Used

9.1.2. The Laboratory Should Have a Documented Approach for Testing General Unknowns

9.1.3. The Procedures Should Include a List of Equipment and Reagents, Step-by-Step Instructions, Quality Controls, Test Calculations, Limitations, Interpretation Criteria, and Literature References

9.1.4. The Laboratory Should Have a Policy Whereby a Deviation from an Analytical Procedure Is Documented and Approved

9.2. The Laboratory Should Use Reagents That Are Suitable for the Methods Employed

9.2.1. The Laboratory Should Have Written Procedures for Documenting Dates of Receipt and Expiration of Commercial Supplies and for the Formulation of Reagents. Lot Numbers Should be Recorded

9.2.2. Reagents Should be Labeled with the Identity of the Reagent, the Date of Preparation and Expiration, the Identity of the Individual Preparing the Reagent, and Storage Instructions

9.2.3. The Laboratory Should Identify Critical Reagents, If Any, and Evaluate Them Prior to Use (e.g., Restriction Enzymes, Primers, Antibodies, Growth Media Prepared in the Laboratory)

10. EQUIPMENT CALIBRATION AND MAINTENANCE

10.1. The Laboratory Should Use Equipment Suitable for the Methods Employed

10.2. The Laboratory Should Identify Critical Equipment and Should Have a Documented Program for Calibration and, When Appropriate, Monitoring of Instruments and Equipment

10.2.1. When Available and Appropriate, Standards Traceable to National or International Standards Should Be Used for Calibration. When Traceability to National Standards of Measurement Is Not Applicable, the Laboratory Should Provide Satisfactory Evidence of Correlation of Results

10.2.2. The Frequency of the Calibration Should Be Documented for Each Instrument Requiring Calibration. Such Documentation Should Be Retained in Accordance with Appropriate Federal and State Law

10.3. The Laboratory Should Have and Follow a Documented Program to Ensure That Instruments and Equipment Are Properly Maintained

10.3.1. New Critical Instruments and Equipment, or Critical Instruments and Equipment That Have Undergone Repair or Maintenance, Should Be Calibrated Before Use

10.3.2. Written Records or Logs Should Be Kept for Maintenance Service Performed on Instruments and Equipment. Such Documentation Should Be Retained in Accordance with Federal and State Law

11. REPORTS

11.1. The Laboratory Should Have and Follow Written Procedures for Generating and Maintaining Documentation for Tested Samples

11.1.1. The Laboratory Should Have Written Procedures for the Release of Tested Sample Information

11.1.2. A Laboratory Report Should Include the Following

- Name of submitting agency
- Date the sample(s) was received in the laboratory
- Brief description of all evidence submitted, including the indicated source of the sample, when available (e.g., lungs, circulatory, water source, debris)
- Statement to address the specific request made of the laboratory
- General statement of analytical method used
- Analytical results and interpretive statement to provide clarity of result
- Statement of negative results, when appropriate
- Statement of recommended additional testing, when appropriate (e.g., additional testing after 6 and 12 days is suggested so as to determine exact exposure date)
- Laboratory location (city, state)
- Identity of the person responsible for the technical conclusions of the laboratory report
- Identification of all components of the analysis completed using subcontractors

11.2. The Laboratory Should Maintain, at a Minimum, the Following Examination Documentation

- Analyst- and technician-generated notes, to include identity of person generating notes

- List of instrumentation used (by specific identifier, if appropriate)
- All charts, graphs, photographs, and other examination-related records
- Chain of custody
- Description of evidence
- Dates of analysis
- Any deviations from examination protocol
- Statement regarding any consumption/disposition of evidence
- The identity of individuals performing technical and administrative reviews of the report

12. REVIEW OF REPORTS

12.1. The Laboratory Should Have and Follow Written Procedures for Technically Reviewing All Tested Sample Information and Results. The Technical Review Is Conducted by a Technically Qualified Individual Other Than the Preparer of the Report, Notes, Data, and Other Documents Under Review. All Reports Should Be Administratively Reviewed

12.1.1. The Laboratory Should Have a Documented Mechanism in Place to Address Discrepant Interpretations and/or Conclusions Between Analyst(s) and Reviewer(s) and a Mechanism to Address Unresolved Conclusions/Interpretations

12.2. The Laboratory Should Have a Practice in Place to Review and Address Reports of Noncompliance (e.g. Findings Resulting from Audit Reports, Contamination Logs, Controls Not Functioning Properly)

13. PROFICIENCY TESTING

13.1. Examiners and Other Personnel Designated by the Technical Manager Who Are Actively Engaged in Analysis Should Undergo Annual Proficiency Testing. The Test Samples Should Be Processed and Analyzed in the Same Manner as Casework. Successful Completion of Annual Testing Is Required to Continue Casework

13.1.1. The Laboratory Should Maintain the Following Proficiency Test Documentation

- Test set identifier
- Identity of the test participant

- Distribution date and due date
- Date of analysis and completion
- All data and notes supporting the conclusions
- Proficiency test results
- Any discrepancies noted
- Corrective actions taken
- Retention, if possible, of a portion of each proficiency test for reanalysis and comparison, if circumstances dictate

Such documentation should be retained in accordance with appropriate federal and state law.

13.1.2. Records for Each Individual Tested Should Be Maintained, Including All Notes, Records, Reports, and Evaluations

13.2. The Following Proficiency Test Information Should Be Documented for Tests Prepared Within the Laboratory

- Identity of person(s) who prepared the test samples
- Date test samples were prepared
- Test identifier
- Sufficient information to duplicate test preparation
- Expected results, including acceptable limits (if appropriate)

14. CORRECTIVE ACTION

14.1. The Laboratory Should Establish and Follow Procedures for Corrective Action Whenever Analytical Errors and/or Proficiency Testing Discrepancies Are Detected. The Laboratory Should Maintain Documentation for the Corrective Action. Such Documentation Should Be Retained in Accordance with Appropriate Federal or State Law

15. AUDITS

15.1. The Laboratory Should Conduct Audits Annually in Accordance with the Guidelines Outlined Herein

15.1.1. Audit Procedures Should Address, at a Minimum, the Following

- Quality assurance program
- Organization and management
- Personnel
- Facilities and security

- Sample control
- Validation
- Analytical procedures
- Calibration and maintenance
- Proficiency testing
- Corrective action
- Documentation and report writing
- Review of reports
- Safety
- Previous audits

15.1.2. The Laboratory Should Retain All Documentation Pertaining to Audits in Accordance with Applicable Federal and State Laws

15.2. A Second Agency Should Conduct an External Audit Once Every Two Years

16. SAFETY

16.1. The Laboratory Should Have and Follow a Documented Environmental Health and Safety Program

This program should include a documented ongoing safety training program for laboratory personnel and should be in accordance with applicable federal and state laws.

16.2. The Laboratory Should Identify an Individual as the Safety Officer

16.3. Laboratories Employing Biosafety Level 2 or Higher Should Meet Applicable Current CDC (BMBL) and/or USDA Guidelines

17. SUBCONTRACTOR OF ANALYTICAL TESTING

17.1. The Submitting Agency Should Be Notified and Approve If Its Samples Will Be Tested by a Subcontracted Laboratory. The Laboratory Will Be Notified and Approve the Use of a Subcontractor by the Subcontractor Laboratory for Testing of Samples

17.2. A Laboratory Operating Under the Scope of These Guidelines Should Require Certification of Compliance with Those Portions of These Guidelines Appropriate to the Analysis When a Subcontractor Performs Analyses for the Laboratory. A Contracting Laboratory May

Choose to Require Documentation of a Laboratory Following These Guidelines Before Submitting Samples for Analysis

17.2.1. The Laboratory Should Establish and Use Appropriate Review Procedures Based on the Type of Work Being Provided by the Subcontracting Laboratory. This is to Verify the Integrity of the Data Received from the Subcontractor. The Review Includes, but is Not Limited to

- Visual or expert system inspection and evaluation of all results and/or data
- Inclusion of quality control samples
- Annual on-site visits

17.3. Laboratories That Use Specialty Contractors to Fulfill Narrowly Focused Analytical Methods (e.g., Genomic Sequencing) Should Require the Contractor to Provide a Minimum of the Following Documentation

- Copy of the step-by-step analytical procedure used by the laboratory
- Identity of the person responsible for analysis results
- Analytical controls used during analysis
- Identity of all persons involved in the testing of the samples
- Chain of custody

17.4. The Contracting Laboratory should Define a Policy for Notification of Applicable Personnel Changes by the Subcontracting Laboratory

BIBLIOGRAPHY

American Society of Crime Laboratory Directors-Laboratory Accreditation Board (ASCLD/LAB), *ASCLD/LAB Accreditation Manual*, April 2001.

Federal Bureau of Investigation, *Quality Assurance Standards for Forensic DNA Testing Laboratories* (1998).

International Standards Organization (ISO)/International Electrotechnical Commission (IEC), *ISO/IEC 17025, General Requirements for the Competence of Testing and Calibration Laboratories* (1999). American National Standards Institute, New York, New York.

U.S. Department of Health and Human Services. *Biosafety in Microbiological and Biomedical Laboratories*, Fourth Edition, U.S. Government Printing Office, Washington, DC, April 1999.

CLIA '88. *Code of Federal Regulations*. Washington, DC, U.S. Government Printing Office, 2001, 57 CFR 7139 pp 883–999.

42 Code of Federal Regulations, Chapter IV (10-1-95 Edition), Health Care Financing Administration, Health and Human Services.

The CDC interim final rule 42 CFR Part 73 Possession, Use, and Transfer of Select Agents and Toxins; Interim Final Rule; FR citation: 67 FR 76886.

The APHIS interim final rule 7 CFR Part 331, repeated at 9 CFR Part 121 Agricultural Bioterrorism Protection Act of 2002; Possession, Use, and Transfer of Biological Agents and Toxins; Interim Final Rule; FR citation: 67 FR 76908.

SUBJECT INDEX

Page numbers followed by *f* indicate figures. Page numbers followed by *t* indicate tables.

A

AAV. *See* Adeno-associated virus
Accelerator mass spectrometry (AMS), 283–286, 284*f*
  for analysis of *Bacillus globigii* spore samples, 286–290, 287*f*
  for analysis of biological signatures, 259–260
  analytical technique, 256*t*
  schematic, 285*f*
Acer, Dr. David, 21
Acquired immune deficiency syndrome (AIDS), 51, 119, 121, 168–169
Acyclovir, for treatment of herpesvirus infections, 46
Adeno-associated virus (AAV), 45
Adult respiratory distress syndrome (ARDS), 147
Aedes, 48
Aflatoxin, 6, 110
AFLP. *See* Amplified fragment length polymorphism
AFM. *See* Atomic force microscopy
African milk bush (*Synadenium compactum* “Ruby”), 4
Agent Orange, 5
Agricultural Research Service (ARS), 17
AIDS. *See* Acquired immune deficiency syndrome
Al Qaeda, 28
AM. *See* Arbuscule mycorrhizal fungi
Amanita phalloides, 110–111
Amanita spp., 110–111
Amanita virosa, 111
Amanitin, 110–111
American Society of Microbiology, 22–23
Amplified fragment length polymorphism (AFLP), 244–245, 254
AMS. *See* Accelerator mass spectrometry
Andrews v. State, 386
Animal and Plant Health Inspection Service (APHIS), 16–17
Animal disease outbreaks, 201–202. *See also* Livestock
  importation of etiologic agents, 229
  surveillance, 193
Annotation, 330–332
  for diagnostics versus traditional genomic annotation, 331–332
Anthrax
  example, 310–311
  serologic analysis of exposure to, 302–306, 304*t*
Anthrax attack (2001), 161, 233, 302, 326
Anthrax spores, 20, 35, 102. *See also Bacillus anthracis*
  clinical presentations, 200
Antibiotics, 110
Antibodies, 255
  monoclonal, 319–320
  tests, 299*t*

Antigens
  response to theoretical, 302*t*
  schematic response, 300*f*
Antiviral drugs, 44
APHIS. *See* Animal and Plant Health Inspection Service
Arbovirus, 81
Arbuscule mycorrhizal (AM) fungi, 114
ARDS. *See* Adult respiratory distress syndrome
Arenaviruses, 57
ARS. *See* Agricultural Research Service
Ascomycota (sac fungi), 114–115
Asocarps, 119*f*
Aspergillus, 110
Aspergillus fumigatus, 121, 122–123
Astroviruses, 47–48, 49
Atomic force microscopy (AFM), 264–270, 265*f,* 266*f,* 267*f,* 268*f,* 269*f,* 270*f*
  for analysis of biological signatures, 258
  analytical technique, 256*t*
  schematic, 267*f*
Attribution, 373
AUDPC, 124
Aum Shinrikyo cult, 3, 28–29, 133

B
B19, 45
Bacillus anthracis, 9, 19, 100, 103–104, 215–216, 217, 241, 242, 244, 245, 262*f*–263*f,* 344
  case study, 357–364
  derivations, 358*f*
Bacillus cereus, 100, 244
Bacillus globigii, 270*f*
  results using accelerator mass spectrometry, 286–290, 287*f*
Bacillus spp., 275*f*
Bacillus thuringiensis, 244, 264
Bacterial pathogens, 99–107. *See also* individual bacteria names
  DNA regions for identification, 99–102
    multi-sequence typing (MLST), 100
    multiple-locus VNTR analysis (MLVA), 101–102
    recently emerged pathogens, 101
    16S rRNA gene sequences, 99–100
  population genetics in a forensic context, 355–369
  viral sequencing strategies, 328
  websites, 103
Baghwan Sri Rajneesh, 3, 28
Baltimore, David, 57–58
BAMS. *See* Bio-Aerosol Mass Spectrometry
Basecalling, 325–327
Basidiomycota, 115, 120*f*
Bergalis v. Acer, 388
BFAC. *See* National Bioforensics Analysis Center
BIG-IV. *See* Botulism immune globulin
Bio-Aerosol Mass Spectrometry (BAMS), 277–279, 279*f*
  for analysis of biological signatures, 259
  analytical technique, 256*t*
Biocrimes. *See* Bioterrorism
Bioforensic Analysis Center, 34, 35
Bioforensics Demonstration and Application Project, 15
Bioinformatics, 19–20, 198–199. *See also* Forensic science
  annotation, 330–332
    for diagnostics versus traditional genomic annotation, 331–332
    tools, 330–331
  bacterial and viral sequencing strategies, 328
  basecalling and accuracy, 325–327
  comparative genomic tools, 336–338
  comparative genomics, 332–333
  definition, 313–314
  diagnostics, 314–315
    chemistries, 317–318
    detection, 315–316
    forensic, 321–323
    nucleic acid detection, 316–317
    protein detection, 319
    protein forensic, 324
  draft versus completed, 327
  genome sequence assembly, 328–330, 329*f*
  genome sequencing and analysis techniques, 324
  high-affinity ligands, 320–321, 320*f*
  large-scale, chip-based techniques, 323–324
  methods for microbial detection and forensic diagnostic design, 313–353
  monoclonal antibody, 319–320
  multisequence and multigenome alignment, 333–336

    pairwise comparison and alignment, 333
    protein structure, 338–339
        analysis tools, 341
        modeling tools, 339–340, 340*f*
    substring comparison, 336
    visualization, 341–342
The Biological and Toxins Weapons Convention (1972), 34, 187–188
Biologic toxins. *See also* individual disease names
    botulinum, 132–146
    Clostridia perfringens epsilon, 148–149
    definition, 131
    deliberate introduction, 159–163
    forensic aspects, 131–156
    forensic handling of threat samples, 213–231
    ricin, 146–147
    Staphylococcal enterotoxin B, 147–148
Biological manufacturing signatures, 251–294
    analytical techniques, 256*t*
    antibodies, 255
    characterization techniques, 258–260
    definition, 252
    detection methodologies advantages and disadvantages, 254*t*
    examples, 253
    fingerprints, 257
    future directions, 290–291
    multidisciplinary approach, 257–258
Biological samples
    disposition, 222–223
    export, 230
    importation of etiologic agents, 228–229
    integrity, 218–219
    legal concerns for sample handling and data records, 219–223
    packaging requirements, 228
    safety issues, 223–225
        biosafety level 1 (BSL-1), 224
        biosafety level 2 (BSL-2), 224–225
        biosafety level 3 (BSL-3), 225
        biosafety level 4 (BSL-4), 225
    shipping regulations, 225–230
    test plans, 213–218
        effects of field conditions on, 217
        effects of field operations on, 216–217
        flow of specimens in analytical laboratory, 214*f*
        forensic handling, 213–231
        standardized analytical assays, 215*f*
    threat credibility, 217–218
    transfer, 228, 229
    transportation, 226–227
Biological warfare
    definition, 28
    means of delivery, 208
Biological weapons
    offensive viewpoint, 31–32
    safety precautions, 204
Biological Weapons Convention, 20
Biological weapons statute (2001), 188
Biosafety, 37–38
Biosafety level 1 (BSL-1), 224
Biosafety level 2 (BSL-2), 224–225
Biosafety level 3 (BSL-3), 225
Biosafety level 4 (BSL-4), 225
Biosecurity, 37–38
    risk assessment, 38
Bioterrorism
    botulism, 141, 143
    credibility, 217–218
    definition, 2, 373
    detection and identification, 10
    educated public and, 21–23
    infectious disease outbreak and, 160*t*
    information and databases, 10–11
    quality assurance, 11
    strain repository, 11
    validation, 11
    work experience and, 208
Bioweapons, 8–9
Bipolaris maydis (Southern corn leaf blight), 121, 123
BLAST, 340
Blue-tongue disease, 4
BoNT. *See* Botulinum neurotoxin
Bornaviruses, 47, 50
Borrelia burgdorferi, 309–310
Botulinum antitoxin, 144–146
Botulinum neurotoxin (BoNT), 3, 5–6, 35, 132–146
    bacteriology, 134
    clinical history, symptoms, and findings, 138, 139*t,* 140
    diagnosis, 143–144
    differential diagnosis, 142*t*
    epidemiology, 140–143
    exposure, 133–134

Botulinum neurotoxin (BoNT) (*continued*)
  intentional, 133–134
  laboratory findings, 143–144
  neurotoxin structure and function, 135–138, 136*f*, 137*f*, 138*f*
  treatment, 144–146
  types, 133
Botulinum toxoid vaccine, 145–146
Botulism. *See* Botulinum neurotoxin
Botulism immune globulin (BIG-IV), 145–146
Bread molds (zygomycota), 114
Brown spot rice, 5
Brucella spp., 105, 218
Brucella abortus, 105
Brucella melitensis, 105
Brucella suis, 105
BSL. *See* Biosafety level
Bunyaviridae (Rift valley fever), 4, 76, 308
Bunyaviruses, 47, 51
Burkholderia mallei (Glanders), 4, 36–37, 102, 105–106, 308
  transmission, 159
  uses, 4
Burkholderia pseudomallei, 105–106, 307–308
Burkitt's lymphoma, 46

C

Caliciviruses, 47–48, 49
California Department of Health Services (CDHS), 145–146
CAMS. *See* Center for Accelerator Mass Spectrometry
Candida spp., 121, 125
Carbon nanotubes (CNTs), 270
Case examples
  analysis, 195
  anthrax spores, 20
  of *Bacillus anthracis,* 357–364
  foot and mouth disease, 175–177
  HIV in dental patients, 21, 29
CASP. *See* Critical Assessment of Structure Prediction
Castor bean *(Ricinus communis),* 6
CDC. *See* Centers for Disease Control
CDHS. *See* California Department of Health Services
Cellulases, 110
Center for Accelerator Mass Spectrometry (CAMS), 283–286
Centers for Disease Control (CDC), 16–17, 22, 51, 76, 103, 132, 192–193, 225–226, 230, 387
  biological sample regulations, 226
  epidemiologic clues, 161–163, 162*t*
  traceback investigation of monkeypox outbreak, 89*f*
Central Intelligence Agency (CIA), role in infectious disease, 32–33
Central nervous system (CNS), 48, 66
Cephalosporin, 110
Cepheid, 318, 349
Cercospora spp., 121, 122
Chain-of-custody (COC), 220
  definition, 373
Chemistry, as bioinformatics diagnostic tool, 317–318
Chestnut blight *(Cryphonectria parasitica),* 123
Chicken pox, 46
Chlostridium perfringens, 243
Chytridiomycota (chytrids), 113–114
CIA. *See* Central Intelligence Agency
CLIA. *See* Clinical Laboratory Improvement Amendments legislation
Climate, disease outbreaks and, 206
Clinical Laboratory Improvement Amendments legislation (CLIA-1988), 372
Clostridia perfringens epsilon toxin, 148–149, 263*f*
ClustalW, 335–336
CNS. *See* Central nervous system
CNTs. *See* Carbon nanotubes
CO92, 104
COC. *See* Chain-of-custody
Coccidioides immitis, 125
CODIS. *See* Combined DNA Index System
Coffee leaf rust, 5
Combined DNA Index System (CODIS), 243
Contigs, 329
Cooperative Agricultural Pest Survey, 193–194
Coronaviridae, 61, 85–86
Coronaviruses, 47–48, 49
Coxiella burnetii (Q fever), 3, 4
Coxsackieviruses, 48
CPE. *See* Cytopathic effect
Crime. *See also* Bioterrorism
  identification, 1

Critical Assessment of Structure Prediction (CASP), 339–340
Crops
  diseases affecting, 5–6
  investigation of disease outbreaks, 202–203
  trade restrictions, 5–6
Croup, 49
Cryphonectria parasitica (chestnut blight), 123
Cryptococcus neoformans, 121, 125
Culex, 48
Cyclosporin, 110
Cytopathic effect (CPE), 69

D

Dangerous goods regulations (DGR), 227–228
Daubert v. Merrell Dow Pharmaceuticals, 382–383
  decision
    analysis, 383
    peer review, 384
    testability, 383–384
  error rate/standards, 384
  general acceptance, 384–386
DDBJ (Japan), 126
Deletion/insertion polymorphism (DIP), 174
Dendograms, 240
Dengue fever virus, 48, 206
Department of Agriculture, 27
Department of Commerce, 230
Department of Defense, 12, 27
  role in infectious disease, 32–33
Department of Energy, 12
Department of Health and Human Services (HHS), 12, 27, 36
Department of Homeland Security (DHS), 11–12
  role in infectious disease, 34
Department of Justice, role in infectious disease, 33
Department of State, role in infectious disease, 34–36
Department of Transportation (DOT), 228
DGR. *See* Dangerous goods regulations
DHS. *See* Department of Homeland Security
DIALIGN, 334
DIP. *See* Deletion/insertion polymorphism
Disease
  animal disease outbreaks investigation, 201–202
  biocrime experience and, 208
  crop disease outbreaks investigation, 202–203
  definition, 109
  DNA forensics of humans and bacteria, 356–357
  document availability, 207
  economic analysis, 206–207
  environmental sampling, 203
  epidemiologic analysis, 195–200
    geographical distribution, 197–199
    global information systems tools, 199
    interviews, 199–200
    with regard to person, animal, or crop affected, 196
  epidemiologic investigation, 194
  equipment delivery, 208
  etiologic agents investigation, 228–229
  evaluation
    of clinical findings, 200–201
    of the etiologic agent, 205–206
    of scientific literature, 207–208
  global reporting systems, 191–192
  investigation steps, 160*t*
  laboratory analysis, 204–205
  mapping, 198
  media reports, 207
  meteorological and climatic conditions analysis, 206
  natural versus deliberate, 187–190
  outbreak investigation, 159, 194–195
  plant epidemiology, 124
  sample handling, 204
  specimen collection, 203
  specimen storage, 204
  surveillance, 158, 190–191
  suspicious outbreaks investigation, 187–212
  transfer of biological agents, 229
  transmission, 148–159, 158*t*
    of AIDS, 168–169
  U.S. reporting systems, 192–194
DNA, 166
  admissibility as evidence, 386–389
  for detection and identification of fungi, 125–128, 126*f*
  engineered novel viruses, 92–93
  as evidence, 7–8
  forensics of humans and bacteria, 356–357
  human mitochondrial, 388–389

DNA (*continued*)
  for identification of bacterial pathogens, 99–102
  pipeline diagram, 346*f*
  typing, 17
  typing services, 386
  viruses, 45–47
DOT. *See* Department of Transportation
Dutch elm disease *(Ophiostoma novo-ulmi),* 123

E

E. coli 0157:H7, 14, 168
EA. *See* Early Antigen
Early Antigen (EA), 297–298
EBNA. *See* Epstein Barr Nuclear Acid
Ebola virus, 36, 50, 67, 68*f*
  transmission, 159
EBV. *See* Epstein Barr virus
Education, as tool to combat bioterrorism, 21–23
EDX. *See* Energy dispersive X-ray microanalysis
Electron microscopy (EM), 67–69, 68*f*
ELISAs. *See* Enzyme-linked immunosorbent assays
EM. *See* Electron microscopy
EMBL (Europe), 126
Encephalitis, 35
  Japanese, 78–79
  St. Louis virus, 48, 81
  Venezuelan equine, 48–49
Endemic, definition, 158
Endocytosis, 42
Energy dispersive X-ray microanalysis (ECX), for analysis of biological signatures, 258
Energy dispersive X-ray microanalysis (EDX), analytical technique, 256*t*
Environment, sampling, 203
Enzyme-linked immunosorbent assays (ELISAs), 43, 71
Epidemic, definition, 158
Epidemiology, 16–17. *See also* Molecular epidemiology
  clues, 161–163, 162*t*
  clusters, 168
  definition, 157–158
  disease analysis, 195–200
  epidemiologic analysis, of disease patterns, 196, 197*f*
  forensic, 194
  of fungi, 123–125
  genotypic methods used in investigations, 167*t*
  geographical distribution of disease, 197–199
  of microorganisms, 235–236
Epstein Barr Nuclear Acid (EBNA), 298
  antibody tests for, 299*t*
Epstein Barr virus (EBV), 46
  schematic response, 300*f*
Equine E antitoxin, 145
Ergot alkaloids, 110
Escherichia coli, 242
Eukaryotic cell, 42, 236
Evidence
  collection, 7
  definition, 7
  DNA, 7–8
  DNA admissibility litigation, 386–389
  legal admissibility, 382–383
  standards for, 381–392
Experts on the Biological Weapons Convention Treaty (1987), 188

F

Famine
  creation, 5
  potato blight in Ireland, 5
FASTA3, 340
FBI. *See* Federal Bureau of Investigation
FDA. *See* U.S. Food and Drug Administration
Federal Bureau of Investigation (FBI), role in infectious disease, 33
Fifth disease, 45
Filoviridae, 50, 61
Flaviviruses, 47–48, 48, 74*f*
Fleas, as weapon, 3
Fluorescent treponemal antibody (FTA), 309
FMD. *See* Foot and mouth disease
Food-borne disease, ribotyping, 243
Foodborne botulism, symptoms and signs, 139*t*
FoodNet, 194
Foot and mouth disease (FMD), 4, 196, 201
  case study, 175–177
  media reports, 207
  virus, 175

Forensic science. *See also* Bioinformatics
  aspects of biologic toxins, 131–156
  case studies, 90–92
  definition, 6–8, 252
  genetic analysis of microorganisms, 233–249
  host factors, 295–312
  of humans and bacteria, 356–357
  molecular epidemiology to viral infections, 51–53
  techniques, 342–352
Francisella tularensis, 102, 104–105, 206, 218, 245
Frye v. U.S., 382–383, 389
FTA. *See* Fluorescent treponemal antibody
Fumonisins, 110
Fungi. *See also* Mushrooms
  arbuscule mycorrhizal, 114
  ascomycota (sac fungi), 114–115
  asexual spores, 116*f*
  basidiomycota (club and mushroom fungi, rusts, and smuts), 115, 120*f*
  biology and detection of human and plant pathogens, 109–130
  chytridiomycota (chytrids), 113–114
  classification, 111–115
  detection and identification, 125–128, 126*f,* 127*f*
  epidemiology, 123–125
  filamentous, 123
  fruiting structures, 117*f,* 119*f*
  genetic analysis, 245–246
  growth habit and reproduction, 115–118
  history, 110–111, 112
  host–pathogen interactions, 121–122
  investigation of disease outbreaks, 202–203
  nomenclature, 111–115, 113*f*
  pathogenicity, 118–123
  as pathogens, 5–6
  taxonomy, 111–115, 113*f*
  thigmotrophic, 122
  vesicular arbuscule mycorrhizal, 114
  zygomycota (bread molds), 114
Fusarium-contaminated flour, 6
Fusarium venenatum, 110, 121

G

G-proteins, 123
GenBank (USA), 126
Gene sequences
  forensic genetic analysis of microorganisms, 233–249
  16S rRNA, 99–100
  for strain typing, 241
Genetic markers, 19–20
Genetics, 19–20
  bacteria in a forensic context, 355–369
Genomes
  analysis, 17–18
  comparative, 332–333, 336–338
  comparison of multiple draft, 338
  consensus gestalt, 347–349, 347*f*
  draft versus completed analysis, 327
  mutation rates, 365
  reference, 329*f*
  sequencing
    assembly, 328–330, 329*f*
    analysis techniques, 324
Geographical Information Systems (GIS), 199
Geography, disease distribution, 197–199
GeoSentinel. *See* Global Emerging Infections Sentinel Network
German measles, 49
Gigospora, 114
GIS. *See* Geographical Information Systems
Glanders *(Burkholderia mallei),* 4, 105–106
  uses, 4
Global Emerging Infections Sentinel Network (GeoSentinel), 192
Global reporting systems, 191–192
  websites, 191–192
Glomus, 114
Government agencies. *See also* individual government agency name
  role in infectious diseases, 32–36

H

HALS. *See* High-affinity ligands
Hand-held nucleic acid analyzer (HANNA), 317
HANNA. *See* Hand-held nucleic acid analyzer
Hanta virus, 36, 76, 203. *See also* Rodents
HBV. *See* Hepatitis B virus
Health, infectious diseases and, 29–31
Hemadsorption, 43
Hemagglutination assay, 70
Hemorrhagic fever, 35
  with renal syndrome (HFRS), 76

Hendra virus, 35, 36, 79–80
Hepadnavirus family, 46, 59
Hepatitis A virus, 48, 175
Hepatitis B virus (HBV), 46
Herpangina, 48
Herpesvirus, 6, 45–46
HFRS. *See* Hemorrhagic fever with renal syndrome
HHS. *See* Department of Health and Human Services
HHV8 (Kaposi's Sarcoma Herpes Virus), 46
Hidden Markov model (HMM), 330–331
High-affinity ligands (HALS), 319, 320–321, 320*f,* 341
HIV. *See* Human immunodeficiency virus
HMM. *See* Hidden Markov model
Host factors, 295–312
HPV16, 46
HPV18, 46
Human cytomegalovirus, 46
Human immunodeficiency virus (HIV)
  in dental patients, 21, 387–388
  as possible bioterrorism attack, 309
  tracking infections, 52
  transmission, 168–169
Human virology, 41–53. *See also* Pathogenic agents; Virus; individual virus names
  classification, 44
  definition, 41–42
  DNA viruses, 45–47
  engineering novel viruses, 92–94
    before/after recombinant DNA, 92–93
    mousepox, 93–94
    source determination, 94
    synthetic poliovirus, 93
  forensic case studies, 90–92
  generation of viral diversity, 62–63
  history, 41
  molecular forensics and epidemiology to viral infections, 51–53
  naturally emerging viruses, 76–89
    monkeypox, 87–89, 87*f,* 89*f*
    Nipah, 78–80
    SARS, 82–87
    sin nombre, 76–78
    West Nile, 80–82, 83*f*–84*f*
  RNA viruses, 47
    negative-sense single-stranded RNA viruses, 49–50
    positive-sense RNA viruses, 47–49
    segmented RNA viruses and the Retroviridae, 50–51
  vaccines, 44
  viral kinetics and outcome, 63–66
    course of infection, 65–66
    immune response, 65
    incubation period, 63–65, 64*f*
  virus anatomy, 56–57
  virus assembly and release, 59–60
  virus classification, 60
  virus family/subfamily, 61
  virus genus, 61–62
  virus identification, 66–75
    complement fixation, 70–71
    cytopathic effect, 69
    electron microscopy, 67–69, 68*f*
    enzyme-linked immunosorbent assay, 71
    hemagglutination assay, 70
    immunostaining, 71
    molecular epidemiology, 72–75, 74*f*
    neutralization assays, 70
    nucleic acid-based methods, 71–72
    serology, 69–70
    sites for virus isolation, 66–67
    virus isolation, 67
  virus life cycle, 57
  virus order, 60–61
  virus replication strategies, 57–59
  virus species, 61–62

I

IATA. *See* International Air Transport Association
ICAO. *See* International Civil Aviation Organization
ICTV. *See* International Committee on the Taxonomy of Viruses
IMMS. *See* Interpolated Markov models
Immune system, 65
Immunoglobulin, classes and properties, 297*t*
Immunostaining, 71
Infant Botulism Treatment and Prevention Program, 145–146
Infectious diseases, 27–39
  biosafety and biosecurity, 37–38
  consequences, 30*f*
  considerations, 306–308
  health viewpoint, 29–31

- offensive biological weapons viewpoint, 31–32
- role of government agencies, 32–36
- routes, 306
- threats and, 36–37

Influenza virus, 51, 192, 206
- repetitive exposure, 301
- response to theoretical antigens from, 302*t*

The Institute for Genomic Research, 103
International Air Transport Association (IATA), 227–228, 228
International Animal Health Code, 190–191
International Civil Aviation Organization (ICAO), 227–228
International Committee on the Taxonomy of Viruses (ICTV), 44
International Office of Epizootics (OIE), 190–191. *See also* World Health Organization
International Organization of Epizootics (OIE), 4
International Society of Travel Medicine, 192
International Standards Organization (ISO), 13
Interpolated Markov models (IMMs), 330–331
Interviews, in suspicious disease outbreaks, 199–200
Iowa Department of Public Health, 103
Irish potato famine *(Phytophthora infestans),* 123
ISO. *See* International Standards Organization

J

Japanese encephalitis virus (JE), 78–79
JE. *See* Japanese encephalitis virus

K

Kaposi's Sarcoma Herpes Virus (HHV8), 46
Kim, 104
Knowledge Center, 12
Kostov, Vladimir, 6
Kumho Tire Co., Ltd. v. Carmichael, 385

L

Laboratory
- accreditation, 374–375
- competency assessment of employees, 376–377
- effects of field operations on, 216–217
- flow of specimens chart, 214*f*
- forensic handling of biological threat samples, 213–231
- proficiency testing, 375–376
- quality assurance monitoring, 376
- quality control testing, 376
- quality management in, 371–379
- records, 378
- reports, 377–378
- security, 378
- standardized analytical assays, 215*f*
- validation of tests, 375

Laboratory Information Management System (LIMS), 221
Laboratory Response Network to Bioterrorism (LRN), 223, 350
Laccases, 110
Lassa fever virus, 51
Lawrence Livermore National Laboratory (LLNL), 15, 318
LIMS. *See* Laboratory Information Management System
Listeria monocytogenes, 168, 242, 243
Livestock. *See also* Animal disease outbreaks
- diseases, 4
- Foot and mouth disease, 4
- importation of etiologic agents, 229
- USDA food regulation, 4

LLNL. *See* Lawrence Livermore National Laboratory
Locard's Exchange Principle, 253
Los Alamos National Laboratory, 15
Lovastatin, 110
LRN. *See* Laboratory Response Network to Bioterrorism
Luminex, 318

M

Magaletti v. State, 389
Magnetic resonance imaging, 339
Malaria, 206
Markov, Georgia, 6
Mastomys coucha, 307
Mastomys natalensis, 307
Maximal exact matches (MEMs), 336
Maximal inspiratory force (MIF), 144

McVeigh, Timothy, 28
Measles, 165
  history, 3
Media, disease outbreaks and, 207
Melioidosis, 102, 105–106, 307–308
MEMs. *See* Maximal exact matches
Microarray analyses, 17
Microbial forensics, 1–25
  antibody tests, 299*t*
  definition, 8–11, 373
  epidemiology, 16–17
  host factors, 295–312
  immunoglobulin classes and properties, 297*t*
  infection considerations, 306–308
  methods for detection and forensic diagnostic design, 313–353
  molecular epidemiology to viral infections, 51–53
  program success, 11–12
  quality management in laboratories, 371–379
  response to theoretical antigens, 302*t*
  schematic response of antigens, 300*f*
  serologic analysis of exposure to anthrax, 302–306, 304*t*
  tools, 17–19
  victims versus perpetrators, 309–311
Microorganisms
  dendograms and phylogenetic trees, 240
  epidemiology, 235–236
  forensic genetic analysis, 233–249
  genetic considerations, 236–237
  molecular genetic techniques for strain typing, 241
  multilocus sequence typing, 241–242
  PCR-based genetic typing, 244–245
  restriction endonucleases and polymerase chain reaction, 237–240, 239*f*
  restriction fragment length polymorphism typing, 242–244
  taxonomy, 235–236
  in viruses and fungi, 245–246
MIF. *See* Maximal inspiratory force
MLEE. *See* Multilocus enzyme electrophoresis
MLST. *See* Multi-sequence typing; Multilocus sequence typing
MLVA. *See* Multiple-locus VNTR analysis
MMWR. *See* Morbidity and Mortality Weekly Report
Molecular epidemiology, 72–75, 74*f*. *See also* Epidemiology
  analysis of genomic sequences, 178–180
  characterization of genetic diversity, 166*t*
  fungal, 127–128, 127*f*
  of RNA viruses, 173–185
  strain typing, 163–169, 165*t*
Monkeypox, 46–47, 87–89, 87*f*, 89*f*
Monoclonal antibody, 319–320
Mononegavirales, 60–61
Morbidity and Mortality Weekly Report (MMWR), 192, 306
Mosquitoes, 66
Mousepox, 93–94
Molluscum contagiosum, 46
Multi-sequence typing (MLST), 100, 127–128, 127*f*
Multilocus enzyme electrophoresis (MLEE), 165, 166
Multilocus sequence typing (MLST), 241–242
Multinational studies, 102
Multiple-locus VNTR analysis (MLVA), 101–102, 244
MUMmer, 334–335, 337, 342, 345
Mushrooms. *See also* Fungi
  as weapon, 3
Mycetozoans, 111
Mycotoxins, 6, 110–111
Myocarditis, 48

N
NACCESS, 350
NAPIS. *See* National Agricultural Pest Information System
NAPs. *See* Neurotoxin-associated proteins
National Agricultural pest Information System (NAPIS), 194
National Biodefense Analysis and Countermeasures Center (NBACC), 11–12
National Bioforensic Evidence Database, 15
National Bioforensic Information Encyclopedia, 15
National Bioforensics Analysis Center (BFAC), 11–12
National Center for Biologic Information, 103
National Institutes of Health (NIH), biological sample regulations, 226

National Notifiable Diseases Surveillance System (NNDSS), 192
National Plant Diagnostic Network, 194
National Research Council, 386
National Science Foundation, 12, 243
NBACC. *See* National Biodefense Analysis and Countermeasures Center
Necrotrophs, 121
Negri bodies, 43
Neisseria meningitidis, 242
Neocallimastix, 113–114
Neurotoxin-associated proteins (NAPs), 135
Newcastle disease, 4, 201
Nidovirales, 60–61
NIH. *See* National Institutes of Health
911 operators, 16
Nipah virus, 35, 36, 78–80
NNDSS. *See* National Notifiable Diseases Surveillance System
Northern Arizona University, 15
Nuclear microscopy. *See also* Particle-induced X-ray emission; Scanning transmission ion microscopy
  for analysis of biological signatures, 260
  analytical technique, 256*t*
Nucleic acid, detection, 316–317

O

OIE. *See* International Organization of Epizootics
Ophiostoma novo-ulmi (Dutch elm disease), 123
ORF. *See* Single open reading frame
Organic acids, 110
Organism, identification, 15
Orthomyxoviruses, 47

P

Palo Verde tree pods, 386–387, 390
Pan Asia O, 176–177
Pandemic, definition, 158
Papillomavirus, 66–67
Papovaviruses, 46
Paramyxoviridae, 49, 61
Particle-induced X-ray emission (PIXE), 281–283, 283*f*
  for analysis of biological signatures, 260
  analytical technique, 256*t*
Parvoviruses, 45
PATHFINDER, 207
Pathogenic agents. *See also* Human virology
  bacterial, 99–107. *See also* Bacterial pathogens
  foodborne, 168
  fungal, biology and detection of humans and plants, 109–130
  history, 3–6
  protein, 351*f*
  recently emerged, 101
  zoonotic, 4, 35, 104
Pathogenicity array analyses, 17
PCR. *See* Polymerase chain reaction inhibitors
Penicillin, 110
Penicillium, 121
People v. Barney, 386
PFGE. *See* Pulsed-field gel electrophoresis
PH. *See* Prospect Hill
Phakospora pachyrhizi (soybean rust), 5
Pharmaceuticals, fungi and, 110
PHLIS. *See* Public health Laboratory Information System
PHS. *See* Public Health Service
Phylogenetic trees, 240
Phytophthora infestans (Irish potato famine), 123
Picornaviruses, 47, 48, 175
PIXE. *See* Particle-induced X-ray emission
Plague, 35, 102. *See also Francisella tularensis; Yersinia pestis*
  clinical presentations, 200
  as weapon, 3
Plant pests, importation, 229
Plant Protection Act (1912), 113
Pleurodynia, 48
Plum Island, 12
Plum Island Animal Disease Center, 38
Pneumonic tularemia, 159
Poisoning
  food, 148
  inhalation, 147
Poliomyelitis, 48, 68*f*
Poliovirus, 48
Polymerase chain reaction (PCR) inhibitors, 17, 237–240, 239*f*
  for detection and identification of fungi, 125–128
  of fungal pathogens, 109
  human virology and, 43

Polymorphism, 244–245, 330
Population studies, 102, 174
  genetics of bacteria in a forensic context, 355–369
Poultry, importation of etiologic agents, 229
Poxviruses, 45, 46–47, 66–67
Proficiency testing, 373, 405
Prospect Hill (PH), 76
Proteases, 110
Protein
  analysis tools, 341
  detection, 319
  forensic, 324
  modeling tools, 339–340, 340*f*
  pathogen, 351*f*
  structure, 338–339
Public health
  considerations for notifying law enforcement, 164*t*
  involvement in disease outbreak investigation, 162–163, 163*t*
Public Health Laboratory Information System (PHLIS), 192–193
Public Health Service (PHS), 228
Puccinia granimis (wheat stem rust), 5, 123
Pulmonary syndrome, 51
Pulsed-field gel electrophoresis (PFGE), 243
PulseNet, 16–17, 167–168, 194
PUU. *See* Puumula
Puumula (PUU), 76
pXO1, 103
pXO2, 103

Q

Q fever *(Coxiella burnetii),* 3, 4
QA. *See* Quality assurance
Quality Assurance Guidelines for Laboratories Performing Microbial Forensic Work, 393–409. *See also* Scientific Working Group of DNA Analysis Methods
  analytical procedures, 402–403
  audits, 406–407
  corrective action, 406
  definitions, 394–396
  equipment calibration and maintenance, 403–404
  facilities, 400
  organization and management, 397
  personnel, 398–400
  proficiency testing, 405
  quality assurance program, 397
  reports, 404–405
  review of reports, 405
  safety, 407
  sample control, 400–402
  subcontractor of analytical testing, 407–408
  validation, 402
Quality assurance (QA), 397
  criteria development and validation, 13–14
  definition, 373
  in the laboratory, 13
  laboratory monitoring, 376
  for microbial forensic laboratories, 11
Quality control, 374
  laboratory testing, 376
Quality systems, 374
Quarantine, foreign regulations, 228–229
Quorn, 110

R

Raman spectroscopy, 271–276
  for analysis of biological signatures, 258–259
  analytical technique, 256*t*
  confocal microscope, 274*f*
  scattering process, 271*f*
  spectra of individual spores, 275*f*
Random amplification of polymorphic DNA (RAPD), 166, 246
RAPD. *See* Random amplification of polymorphic DNA
Reoviridae, 51, 93
Reoviruses, 57
Respiratory syncytial virus, 49–50
Restriction fragment length polymorphism/pulsed-field gel electrophoresis (RFLP-PFGE), 166, 237–240, 239*f*
Restriction fragment length polymorphism (RFLP), 242–244, 321
Retroviridae, 50–51, 59
RFLP. *See* Restriction fragment length polymorphism
RFLP-PFGE. *See* Restriction fragment length polymorphism/pulsed-field gel electrophoresis

Rhabdoviridae, 50, 61
Rhinovirus, 48, 175
Rhizoctonia solani, 112
Ribosomal Database Project, 243
Ribotyping, 243
Ricin, 6, 146–147
  clinical presentations, 201
  clinical signs and symptoms, 147
  inhalation poisoning, 147
  structure and action, 146–147
  therapy, 147
Ricinus communis (Castor bean), 6
Rift Valley fever *(Bunyaviridae)* virus (RVFV), 4, 36–37, 102–106, 308
  transmission, 159
  uses, 4
RNA viruses, 47
  molecular epidemiology and forensics, 173–185
  negative-sense single-stranded RNA viruses, 49–50
  positive-sense, 47–49
  ribotyping, 243
  segmented viruses and the retroviridae, 50–51
Rodents, 77–78. *See also* Hanta virus
Roseola, 46
rRNA, 243
Rusts, 115, 120*f,* 122
Rye ergot, 6

S

16S rRNA gene sequences, 99–100
Sac fungi (ascomycota), 114–115
Saccharromyces cerevisiae, 110
Salmonella enterica, 243
Salmonella typhimurium, 3, 14, 102–103, 160–161, 168
SARS. *See* Severe acute respiratory syndrome
Scanning electron microscopy (SEM), 261–264, 262*f*–263*f*
  for analysis of biological signatures, 258
  analytical technique, 256*t*
Scanning transmission ion microscopy (STIM), 281–283, 283*f*
  for analysis of biological signatures, 260
  analytical technique, 256*t*
Scientific literature, disease evaluation and, 207–208
Scientific Working Group of DNA Analysis Methods (SWGDAM), 12, 390. *See also Quality Assurance Guidelines for Laboratories Performing Microbial Forensic Work*
Scientific Working Group on Microbial Genetics and Forensics (SWGMGF), 12–15, 371
Sclerotinia sclerotiorum, 121
Scutellospora, 114
SEB. *See* Staphylococcal enterotoxin B
Security, in the laboratory, 378
SEM. *See* Scanning electron microscopy
September 11, 2001, 28
Sequencing, 17
Serology, 69–70
SERS. *See* Surface-enhanced Raman spectroscopy
Severe acute respiratory syndrome (SARS), 49, 68*f,* 82–87, 306–307
  transmission, 159
Sexually transmitted diseases, 46
Shigella dysenteriae, 3, 14, 102–103, 168
Shigellosis, 3
Shingles, 46
Sin nombre virus (SNV), 76–78
Single-nucleotide polymorphism (SNP) characteristics, 17, 174
Single open reading frame (ORF), 175
SLE. *See* St. Louis encephalitis virus
Smallpox, 22, 29–30, 46, 315
  history, 3
  vaccination strategies, 22
Smuts, 115, 120*f*
SNARE. *See* Soluble N-ethylmaleimide-sensitive factor attachment protein receptor
SNP. *See* Single-nucleotide polymorphism characteristics
SNV. *See* Sin nombre virus
Soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE), 135, 137, 137*f,* 138*f*
SOP. *See* Standard operating procedure
Southern corn leaf blight *(Bipolaris maydis),* 121, 123
Soybean rust *(Phakospora pachyrhizi),* 5
Specimens
  collection, 203
  decontamination procedures, 203
  storage, 204

Spores
  analytical techniques, 256*t*
  fungal, 116*f*
  weaponized, 262*f*–263*f*
St. Louis encephalitis virus (SLE), 48, 81
Standard operating procedure (SOP), 377
Staphylococcal enterotoxin B (SEB), 146, 147–148, 205
  clinical signs and symptoms, 148
  structure and action, 147–148
  therapy, 148
Staphylococcus aureus, 242
State of Louisiana v. Richard J. Schmidt, 90–92
State v. Bogan, 386–387
State v. Schmidt, 388
STIM. *See* Scanning transmission ion microscopy
Summary of Notifiable Diseases, United States, 192
Surface-enhanced Raman spectroscopy (SERS), 271–276
  for analysis of biological signatures, 258–259
  analytical technique, 256*t*
  colloidal silver particles, 276*f*
  confocal microscope, 274*f*
  scattering process, 271*f*
  spectra of individual spores, 275*f*
SWGDAM. *See* Scientific Working Group of DNA Analysis Methods
SWGMGF. *See* Scientific Working Group on Microbial Genetics and Forensics
Synadenium compactum “Ruby” (African milk bush), 4
Synchytrium endobioticum, 113
SynPlot, 342
Synthetic poliovirus, 93
Syphilis, 309

T

T-cell leukemia, 51
Tensilon test, 143
Terrorism
  costs, 2
  definition, 2, 28–29
  investigation of suspicious disease outbreaks, 187–212
Test validation, 374, 407–408
Thermus aquaticus, 238
Threats, 36–37
TIGR, 344
Time-of-flight secondary ion mass spectrometry (ToF-SIMS), 279–281
  for analysis of biological signatures, 259
  analytical technique, 256*t*
  schematic, 281*f*
ToF-SIMS. *See* Time-of-flight secondary ion mass spectrometry
Togaviruses, 47–48, 48–49
Toxins, as weapons, 5–6
Trichothecenes, 110
Tularemia, 35, 159, 203
  clinical presentations, 200–201

U

Unexplained Deaths and Critical Illnesses Surveillance System, 193
United Nations Special Commission (UNSCOM), 189
United States Postal System (USPS), 228
United States v. Chischill, 385
United States v. Plaza, 390
UNSCOM. *See* United Nations Special Commission
Untranslated regions (UTRs), 175
Uromyces appendiculatus, 122
U.S. Army Medical Research Institute for Infectious Diseases (USAMRIID), 81, 213, 305
U.S. Department of Agriculture (USDA), 27, 113, 168, 193, 229
  food regulation, 4
U.S. Department of Commerce, 230
U.S. Department of Energy, 12, 243
U.S. Federal Bureau of Investigation, 371
U.S. Food and Drug Administration (FDA), 168, 194
U.S. Geological Survey, 193
U.S. Patriot Act (2001), 188
USAMRIID. *See* U.S. Army Medical Research Institute for Infectious Diseases
USDA. *See* U.S. Department of Agriculture
USPS. *See* United States Postal System
UTRs. *See* Untranslated regions

V

Vaccines
  antibody response and, 297

  for botulism, 145–146
  to control viral infections, 44
  for measles, mumps, rubella, 49
VAM. *See* Vesicular arbuscule mycorrhizal fungi
Variable Number Tandem Repeat (VNTR) loci, 101–102, 321–322
Varicella zoster, 46
Variola, 46
Variola virus, 29–30
VC. *See* Vital capacity
VCA. *See* Viral Capsid Antigen
VECTOR, 22
Venezuelan equine encephalitis, 48–49
Vesicular arbuscule mycorrhizal (VAM) fungi, 114
Vesicular stomatitis, 4
Veterinary services, 193
Vibrio cholerae, 244
Viral Capsid Antigen (VCA), 297–298
Viral kinetics, 63–66
  course of infection, 65–66
  immune response, 65
  incubation period, 63–65, 64*f*
Virus. *See also* Human virology; name of virus
  adaptive immune response, 64*f*
  anatomy, 56–57
  assembly and release, 59–60
  classification, 60
  family/subfamily, 61
  genetic analysis, 245–246
  genus, 61–62
  identification, 66–75
  life cycle, 57
  order, 60–61
  replication strategies, 57–59
  species, 61–62
Vital capacity (VC), 144
Vmatch, 336, 349
VNTR. *See* Variable Number Tandem Repeat loci

W
Websites
  for detection and identification of fungi, 126
  for economic analysis, 206–207
  for foreign quarantine, 229
  on fungi, 110
  global information systems, 199
  for global reporting systems, 191–192
  for handling biological samples, 227
  health departments, 146
  for legislation, 372
  for nucleic acid detection, 317
  for plant disease epidemiology, 124
West Nile virus (WNV), 66–67, 80–82, 83*f*–84*f,* 192, 206, 315
  transmission, 159
Wheat stem rust *(Puccinia granimis),* 5, 123
WHO. *See* World Health Organization
Wildlife, 35–36
WNV. *See* West Nile virus
World Health Organization (WHO), 34, 82–83, 188

X
X-ray crystallography, 339
Xylanases, 110

Y
Yeast, 110
Yellow fever virus, 48
Yellow rain, 6
Yersinia pestis, 102, 104, 205–206, 245, 307, 327, 348
Yersinia pseudotuberculosis, 327, 348

Z
Zearalenone, 110
Zoonotic pathogens, 4, 35, 104, 195
Zoster, 46
Zygomycota (bread molds), 114